**Analysing Patterns and Risk Factors of Gun Violence in the US**

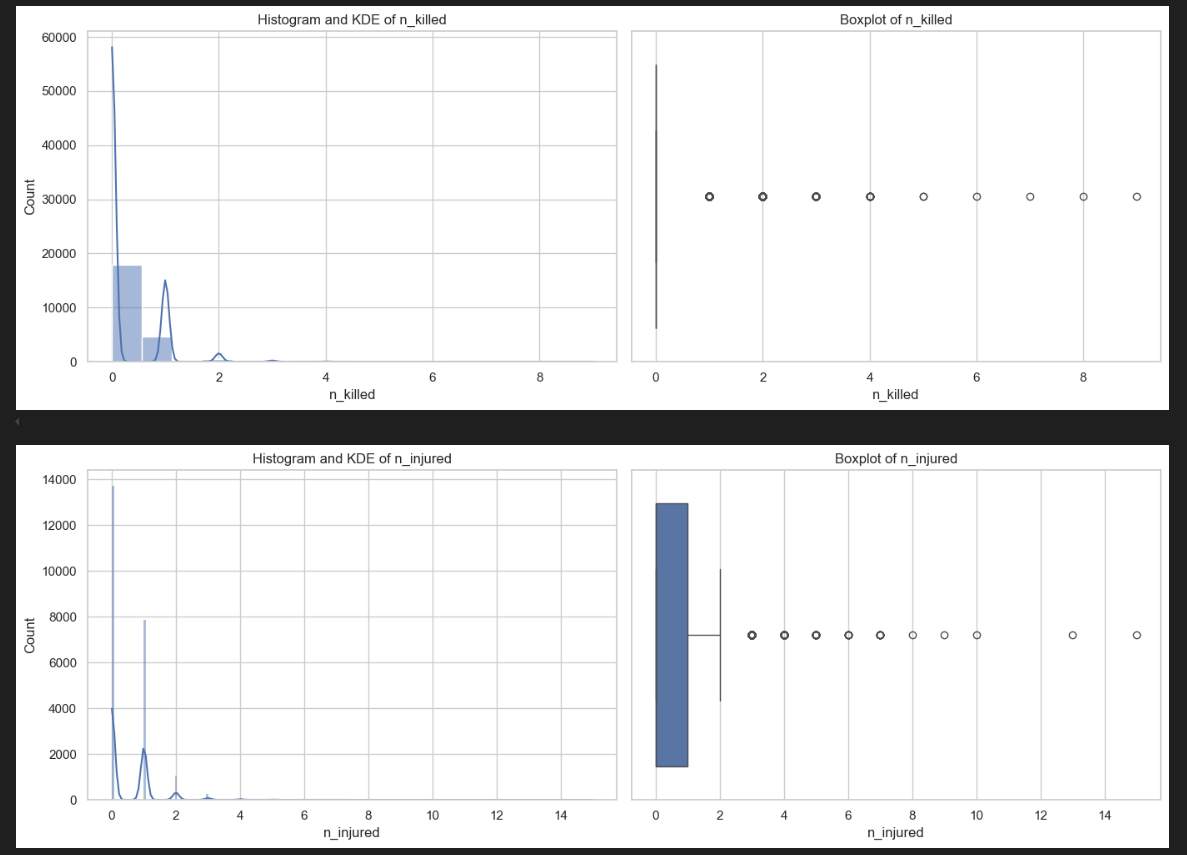
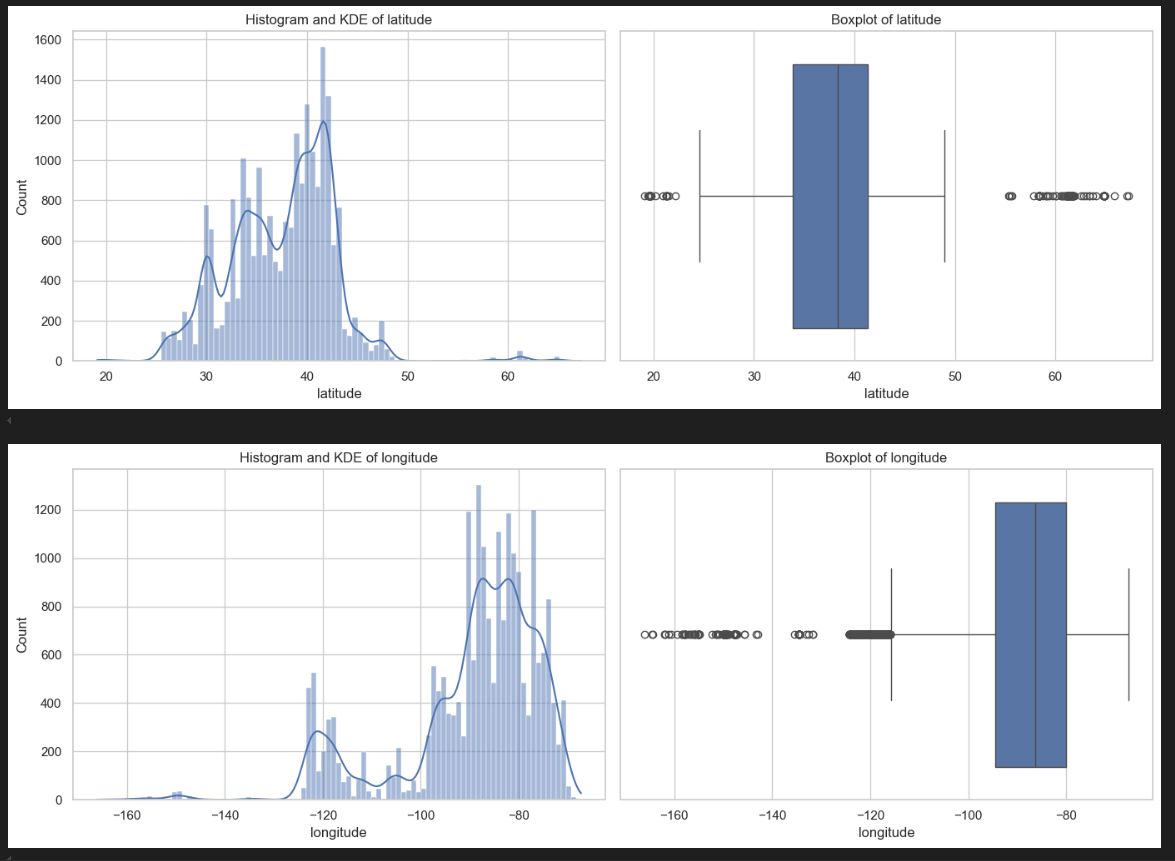
**Data Science Project**

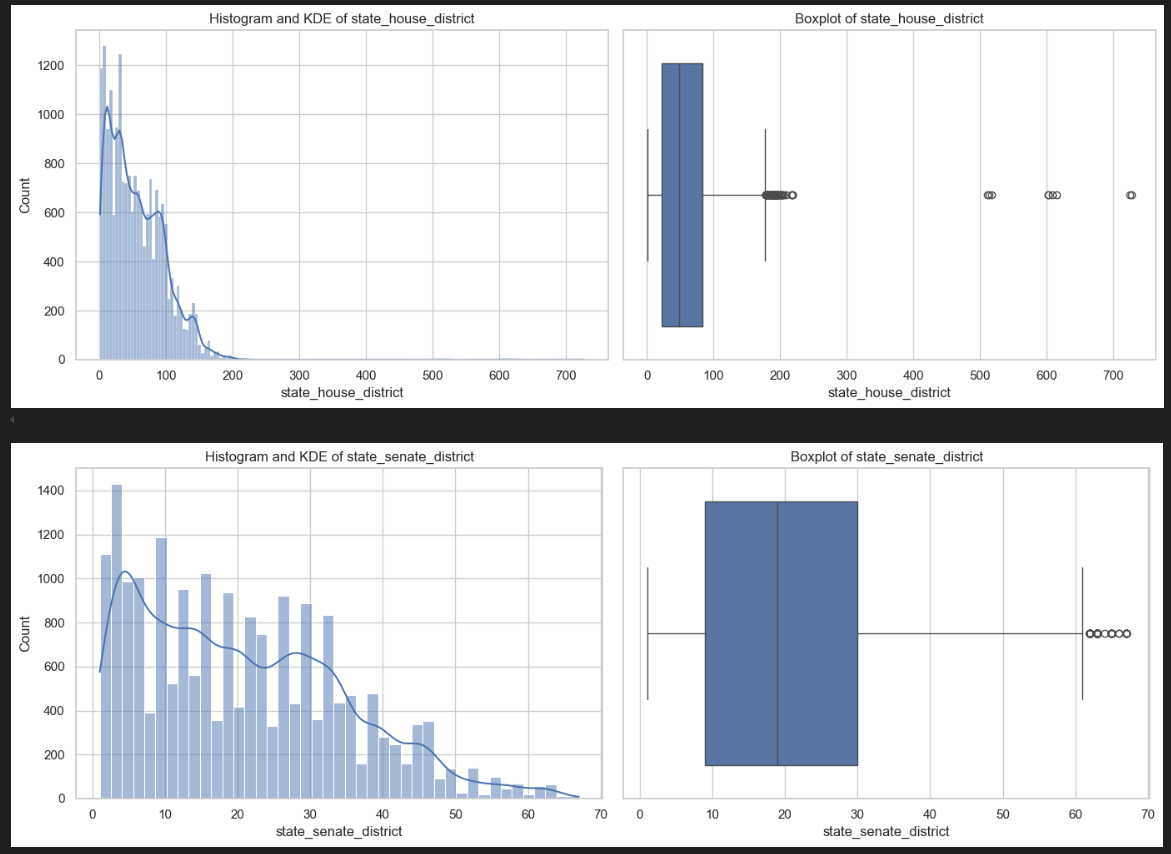
**Exploratory Data Analysis and Visualization Report**

|  |  |
| --- | --- |
| **Name** | **Roll Number** |
| Shaaf Salman | 21L-6083 |
| Syed Farhan Jafri | 21L-6074 |
| Haider Khan | 21L-6067 |

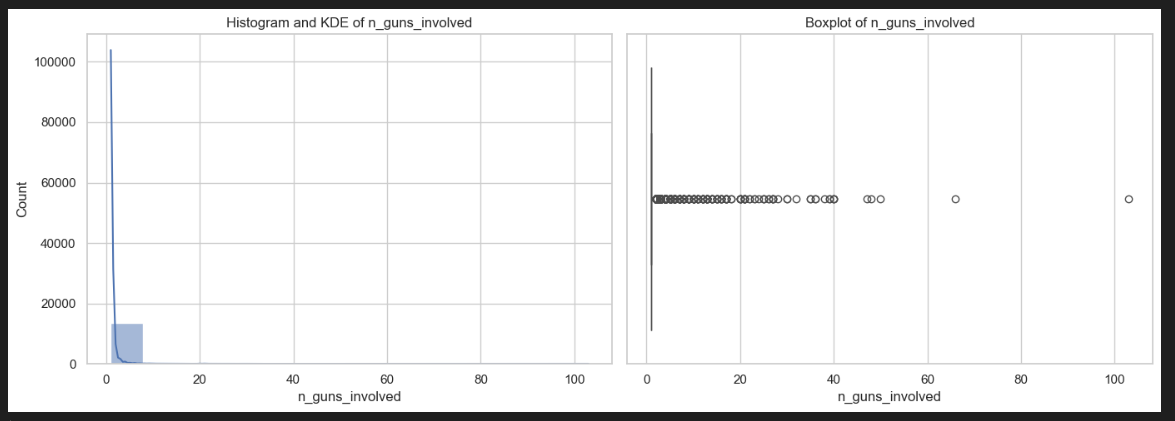
1. **Summary Statistics and Visualizations for Each Variable**

**Numerical Variables**

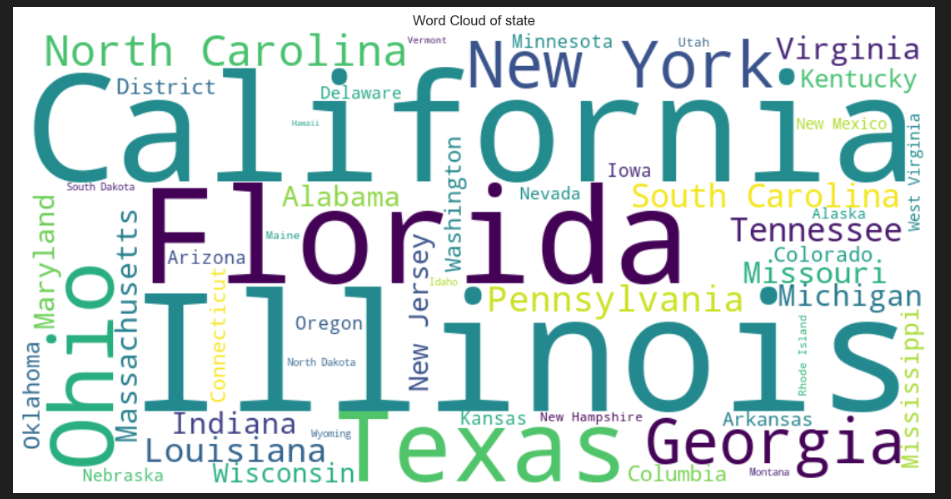
* **n\_killed and n\_injured**: Most incidents involved 0–2 deaths or injuries. Outliers present showing some incidents had 10+ casualties. Histograms show highly right-skewed distributions (most incidents had very few victims). Boxplots confirm the presence of extreme outliers.
* **latitude and longitude:** Values are distributed according to U.S. geography. No major outliers outside the expected U.S. range.

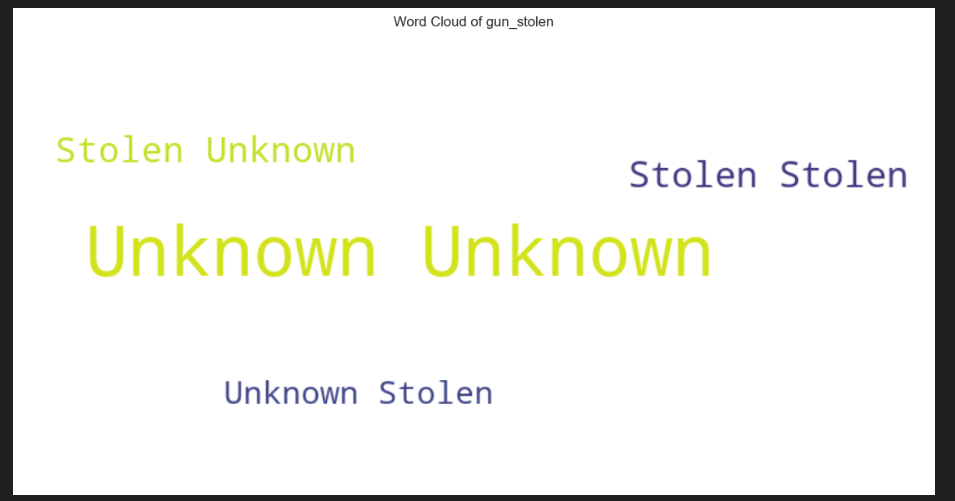


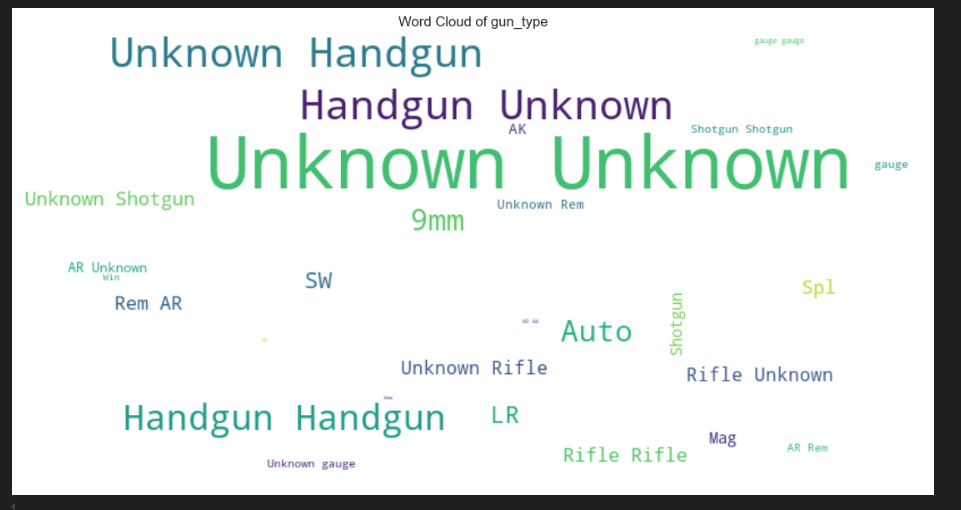
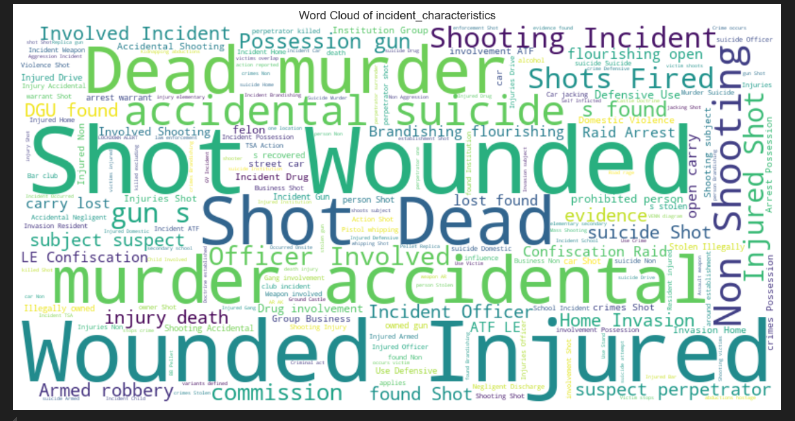
* **state\_house\_district and state\_senate\_district:** Distributed fairly evenly across districts. Some missing values where district information was unavailable.
* **n\_guns\_involved:** Mostly 1–2 guns per incident. Some extreme cases with 10+ guns.



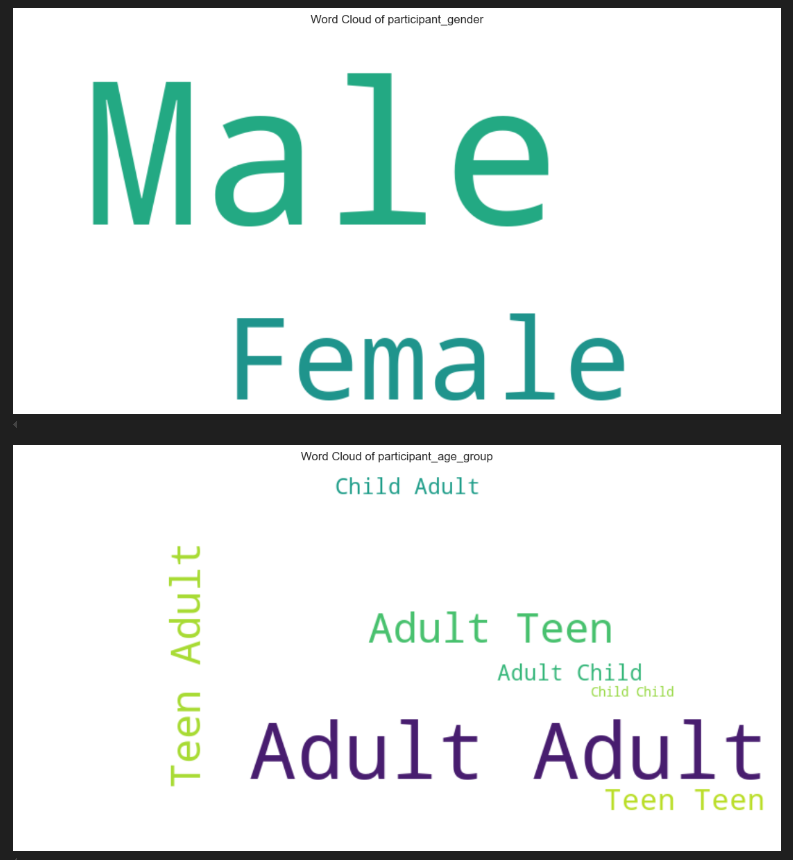
**Categorical Variables**

* **state:** States like Illinois, California, and Texas had the highest number of incidents.
* **gun\_stolen:** Majority of incidents had unknown or unspecified stolen gun status. Among known cases, "Not-stolen" guns were more common than "Stolen".



* **gun\_type:** "Handgun" was the most frequent weapon type used. Word clouds revealed additional types like rifles and shotguns.
* **incident\_characteristics**: Frequent terms included "Home Invasion", "Drive-by", "Argument", indicating common types of gun violence.

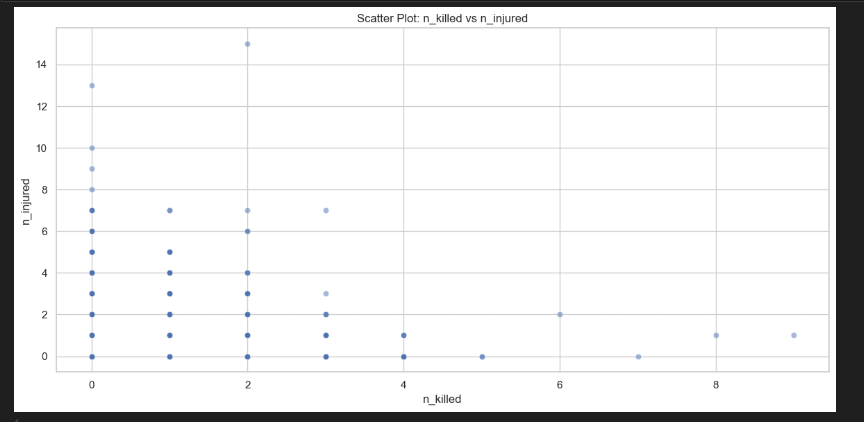
* **participant\_gender** **and** **participant\_age\_group**: Most participants were adult males (18+ years old). Fewer female participants overall.

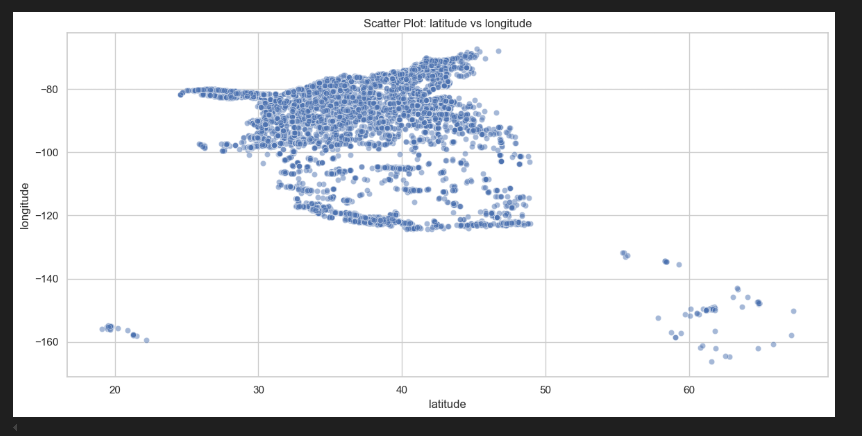


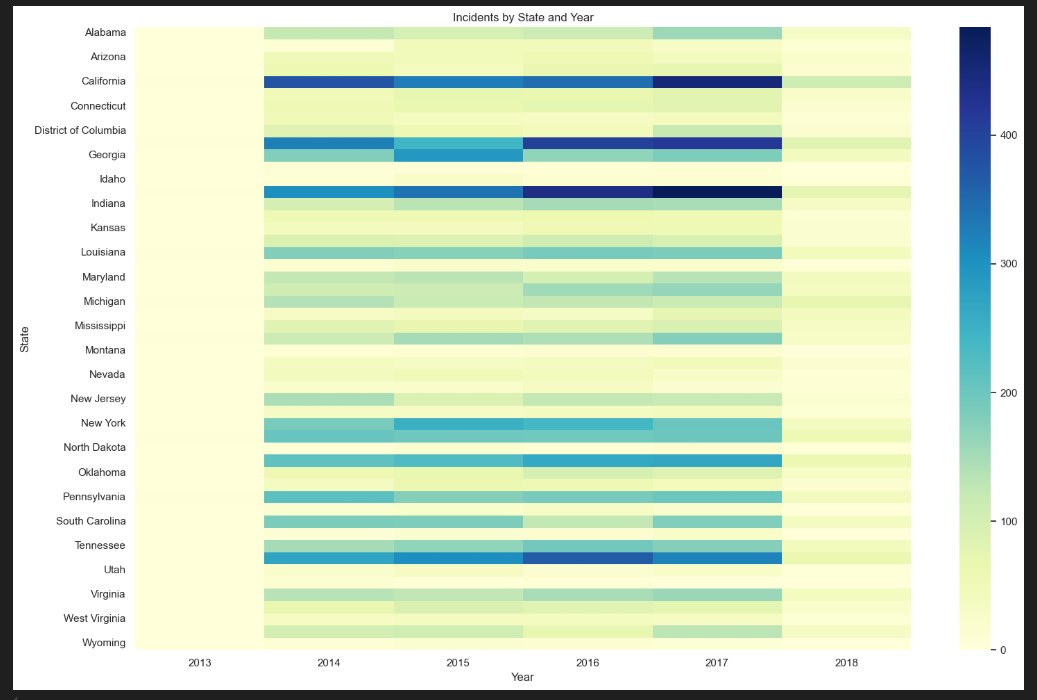
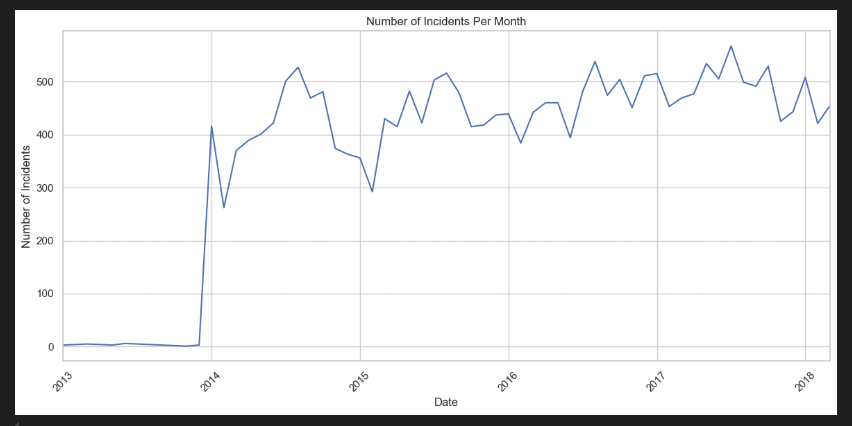
**2. Insights and Observations from Univariate, Bivariate, and Multivariate Analysis**

**Univariate Analysis:** Gun violence incidents often result in either injuries or deaths, but rarely both. Incidents typically involve a small number of participants and few weapons.

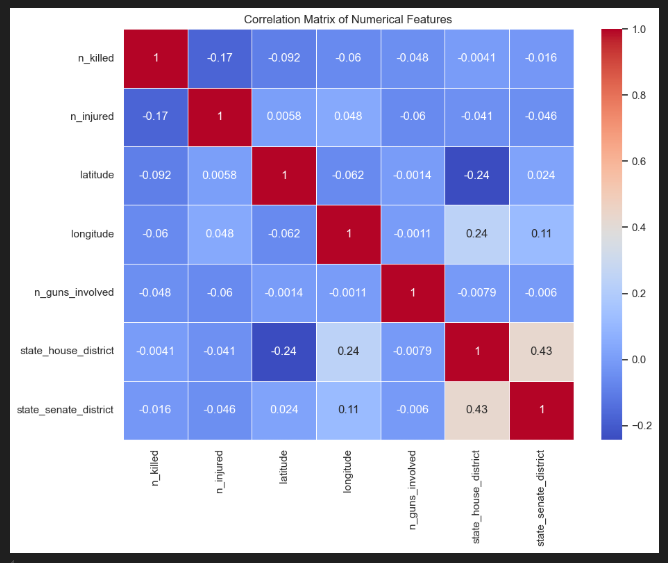
**Bivariate/Multivariate Analysis:**

* **n\_killed vs n\_injured:** Positive correlation as incidents with more injuries also tend to have more deaths, though not perfectly.
* **latitude vs longitude:** Scatter plot shows clustering along major urban areas like the East Coast and Midwest.



* **State vs Year heatmap**: Gun violence is persistent over the years, with some states (like Illinois and California) consistently showing high incident counts.
* **Monthly trend:** Slight seasonal variation visible: incidents peak slightly in the summer months (June–August).

**Correlation Matrix**

****Moderate positive correlation between n\_killed, n\_injured, and n\_guns\_involved. This suggests that incidents involving a higher number of guns tend to result in more casualties, both injuries and deaths. Although the correlation is not extremely strong, it indicates a meaningful relationship where an increase in the number of weapons involved could escalate the severity of an incident.

1. **Key Findings and Insights from the Descriptive Analysis**

* **Concentration**:  
  A small number of states account for a large proportion of incidents.  
  Urban areas are hotspots for gun violence.
* **Casualties**:   
  Most incidents result in few casualties.  
  However, a few extreme incidents (outliers) cause mass casualties.
* **Weapons**:   
  Handguns dominate gun violence incidents.   
  Stolen guns are less frequent but still a significant concern when known.
* **Seasonality**:   
  Gun violence shows slight increases during summer months, suggesting a seasonal pattern.
* **Data Gaps**:   
  Missing values in participant details (age, gender, relationship) and district fields could limit the depth of participant-focused analyses.