Data Visualization HomeWork1.

## TABLE OF CONTENTS




**NARAIN SARATHY JAWAHAR**

**1002223825**

## ***1 INTRO***

I have got the region_6 for my assignment. My dataset contains **54,725** records and **135** columns. It includes information on terrorism attacks, with features such as:

Your dataset on terrorism attacks contains 57 features, capturing various aspects of each incident. Here's a breakdown of the key features:

**Time-Related Features**

• **iyear** – The year of the attack. Helps in analyzing trends over time.

• **imonth** – The month of the attack. Useful for identifying seasonal patterns.

• **iday** – The specific day of the attack.

**Location-Based Features**

• **country_txt** – The country where the attack occurred.

• **region_txt** – The broader region (e.g., Middle East, South Asia, etc.).

• **provstate** – The state or province within the country.

• **city** – The specific city where the attack happened.

• **latitude & longitude** – The exact geographical coordinates of the attack, useful for mapping and geospatial analysis.

**Attack Success & Suicide Indicators**

• **success** – Indicates whether the attack was successful (1) or not (0).

• **suicide** – Indicates whether the attack was a suicide attack (1) or not (0).

**Attack Characteristics**

• **attacktype1_txt** – The primary attack type (e.g., Bombing, Armed Assault, Hijacking).

• **attacktype2_txt** & **attacktype3_txt** – Secondary and tertiary attack types if multiple methods were used.

• **targtype1_txt** – The broad category of the target (e.g., Government, Private Citizens, Military).

• **targsubtype1_txt** – A more specific target category (e.g., Police, Schools, Businesses).

• **targtype2_txt, targsubtype2_txt, targtype3_txt, targsubtype3_txt** – Additional target details if multiple targets were involved.

**Perpetrator Details**

• **gname** – The name of the group responsible for the attack (e.g., Al-Qaeda, ISIS).

• **gname2 & gname3** – Secondary and tertiary perpetrator groups if multiple groups were involved.

• **guncertain1, guncertain2, guncertain3** – Flags indicating uncertainty about the identified perpetrators.

**Casualties & Injuries**

• **nkill** – Number of people killed in the attack.

• **nkillus** – Number of US citizens killed.

• **nkillter** – Number of terrorists killed.

• **nwound** – Number of people injured.

• **nwoundus** – Number of US citizens injured.

• **nwoundte** – Number of terrorists injured.

**Weapon Information**

• **weaptype1_txt** – The main type of weapon used (e.g., Explosives, Firearms, Chemical).

• **weapsubtype1_txt** – More specific details about the weapon (e.g., Car Bomb, AK-47).

• **weaptype2_txt, weapsubtype2_txt, weaptype3_txt, weapsubtype3_txt, weaptype4_txt, weapsubtype4_txt** – Additional weapon details if multiple weapons were used.

**Hostage & Kidnapping Details**

• **ishostkid** – Indicates if hostages were taken (1) or not (0).

• **nhostkid** – Number of hostages taken.

• **nhostkidus** – Number of US hostages.

• **ransom** – Indicates if a ransom was demanded (1) or not (0).

• **ransomamt** – Amount of ransom demanded.

**Motive & Additional Information**

• **motive** – Text describing the possible motive behind the attack.

• **addnotes** – Any additional notes or context provided about the attack.

## ***2 OBJECTIVES***

**Identifying the Most Dangerous Attack Zones**

• Analyze which **countries, regions, states, and cities** have experienced the highest number of attacks.

• Use heatmaps and frequency analysis to pinpoint high-risk areas.

**Examining Weapons Used & Their Impact**

• Identify the most commonly used **weapons** and their subtypes.

• Compare weapon types with **casualty numbers** to determine which are the deadliest.

**Assessing Casualties & Attack Severity**

• Compute the **average fatalities and injuries per attack** to assess the deadliest incidents.

• Identify the most lethal attacks based on **total casualties (nkill + nwound)**.

**Analyzing Terrorist Groups & Their Targets**

• Identify the **top terrorist organizations** responsible for the highest number of attacks.

• Determine the **most targeted places or groups** (e.g., civilians, government, military).

• Compare different groups' **preferred attack methods and weapons**.

**Understanding Yearly Terrorism Trends**

• Analyze **yearly attack trends** to identify patterns and fluctuations over time.

• Use time-series analysis to detect any **spikes or declines in terrorist activity**.

## ***3. Methodology***

### 3.1 Analytical Approach

We employ descriptive statistics, data visualization, and exploratory analysis to extract meaningful insights from the dataset.

### 3.2 Data Transformations

To enhance the quality and reliability of the analysis, several data transformations were applied:

• **Handling Missing Values:**

• Features with more than **50% missing values** were **removed** to maintain data integrity.

• Remaining missing values in critical columns were handled based on their significance:

• **Categorical data** (e.g., targsubtype1_txt, weapsubtype1_txt) was replaced with "Unknown" or "Not Specified".

• **Numerical values** (e.g., nkill, nwound) were imputed with **zero** (assuming no casualties were recorded) or the **median** when appropriate.

• **Feature Selection & Filtering:**

• Features were **filtered based on analysis objectives**, retaining only those relevant to attack locations, weapons used, casualties, terrorist groups, and yearly trends.

• Columns related to **unnecessary metadata or redundant information** were dropped.

• **Feature Engineering:**

• Created a **Total Casualties** column (nkill + nwound) to measure attack severity.

• Generated **yearly and regional aggregates** to observe trends over time.

***4. Python vs. R: Usability & Visualization Quality***

When working with our dataset, both Python and R had their strengths and challenges. Here's how they compared in terms of ease of use and visualization quality.

**Python: Flexible but Requires Effort**

**Ease of Use**

• Python gave us full control over data processing and analysis, making it great for handling large datasets.

• The Pandas library made filtering and cleaning data smooth, but it required more coding compared to R.

• It integrates well with machine learning tools, so if we wanted to go beyond visualization, Python was the way to go.

Downside: Customizing charts, especially in Matplotlib, took extra effort to get them looking right.

**Visualization Quality**

• Seaborn & Matplotlib helped us create bar charts, histograms, and trend lines, but tweaking them for better aesthetics required additional coding.

• Plotly made it easy to create interactive charts, which helped in exploring trends dynamically.

• Folium & Geopandas worked well for mapping attack locations, making it easy to spot high-risk zones.

**Downside:** Some of Python's default charts didn't look great without modifications.

R: Quick & Beautiful Visuals, But Less Interactive

**Ease of Use**

• R's Tidyverse (dplyr, tidyr) made cleaning and summarizing data much faster than Python.

• Its built-in statistical functions were perfect for quick insights without writing too much code.

**Downside:** R wasn't as flexible when handling really large datasets, and it's not the best choice for complex transformations.

**Visualization Quality**

• ggplot2 produced stunning, publication-ready visuals with minimal effort.

• Faceting allowed us to compare different attack patterns side by side easily.

• Heatmaps & Boxplots looked more refined in R than in Python.

**Downside:** Unlike Python's Plotly, R's charts weren't interactive, which made exploring the data a bit more rigid.

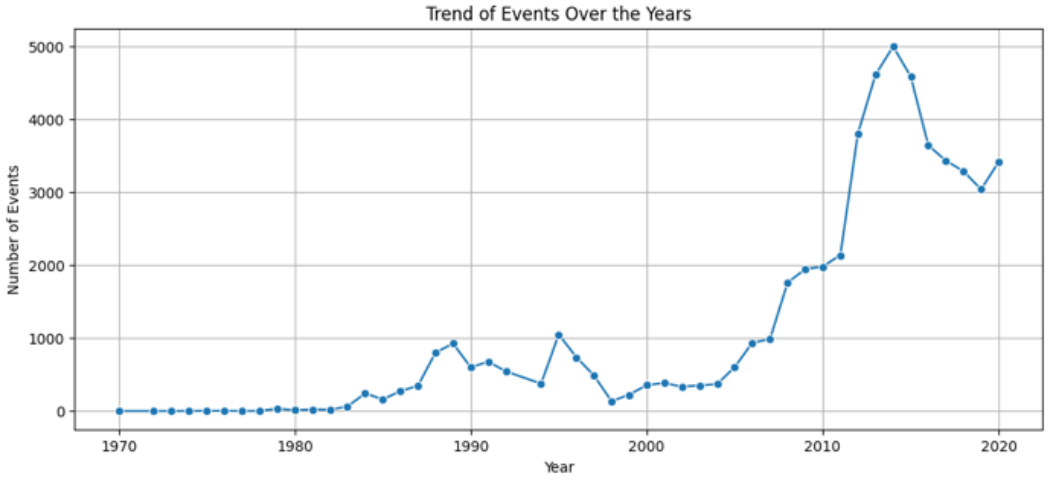
**Final Verdict:**

• Python was better for processing large datasets, customizing visuals, and creating interactive charts.

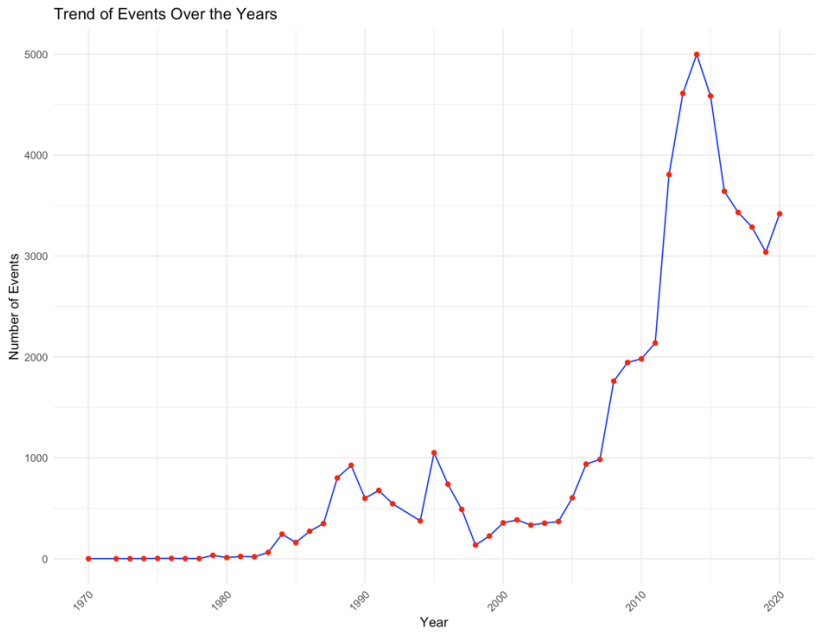
• R was quicker for generating beautiful, polished charts with less effort, but it lacked interactivity.

## *4. Findings:*

## Python-



Data Visualization HomeWork1.



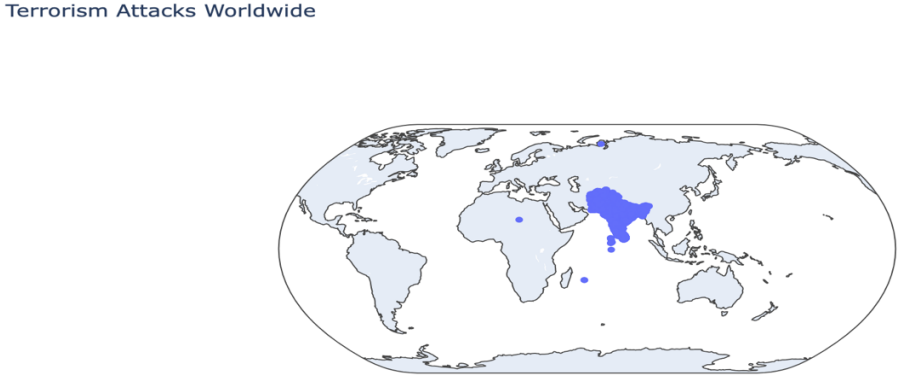

**R STUDIO**

## INFERENCE

The graph shows the trend of terrorism events over the years. From the early 1970s to around 1980, the number of events remained very low and stable. However, starting in the 1980s, there was a gradual increase in incidents, with some fluctuations.

A significant spike is observed in the early 1990s, followed by a decline. The number of events remained relatively moderate until the mid-2000s, after which there was a sharp and consistent rise. Around 2010, incidents increased drastically, peaking between 2014 and 2016, reaching their highest point. After this peak, the numbers started declining but remained high compared to earlier decades.This trend indicates that terrorism incidents have grown substantially over time, with a few periods of decline but an overall upward pattern, especially after 2010.

**GEOSPATIAL PLOTTING OF TERRORISM ATTACKS**:




Python

## INFERENCE

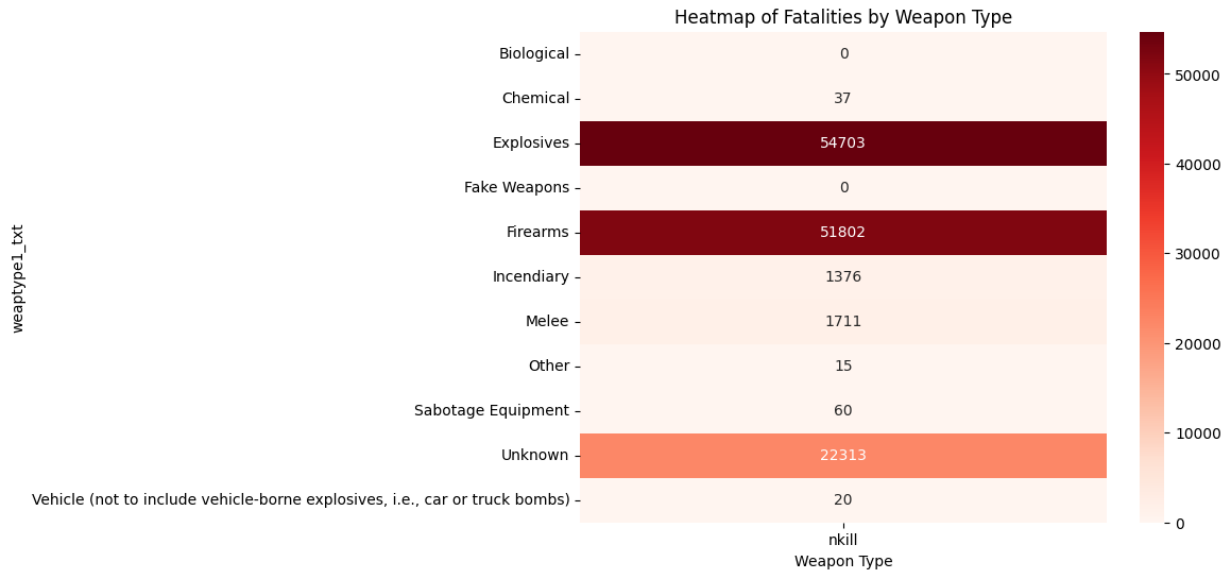
This map provides a global overview of terrorism attacks, highlighting regions with higher concentrations of incidents. The larger and darker blue areas indicate countries that have experienced more attacks, with South Asia, particularly around India and its neighboring countries, showing a significant number of incidents. Other regions, including parts of Africa and isolated locations across the world, also have recorded events, though at a smaller scale.

This visualization helps in understanding where terrorism has been most prevalent, offering insights into geographic patterns and potential hotspots of activity worldwide. I have included other regions as outliers so that I can dive deep into analytics on the terrorism happening in South Asia.

Step 1: Define the South Asia Boundary

South Asia includes India, Pakistan, Bangladesh, Sri Lanka, Nepal, Bhutan, Maldives, and Afghanistan. Approximate latitude and longitude ranges: • Latitude: 0° to 40° (South Asia lies roughly between the equator and 40°N) • Longitude: 60° to 100° (South Asia lies between 60°E and 100°E)

## HEATMAP OF FATALITIES BY WEAPONTYPE AND NKILL




Img provided by python

**INFERENCE:**

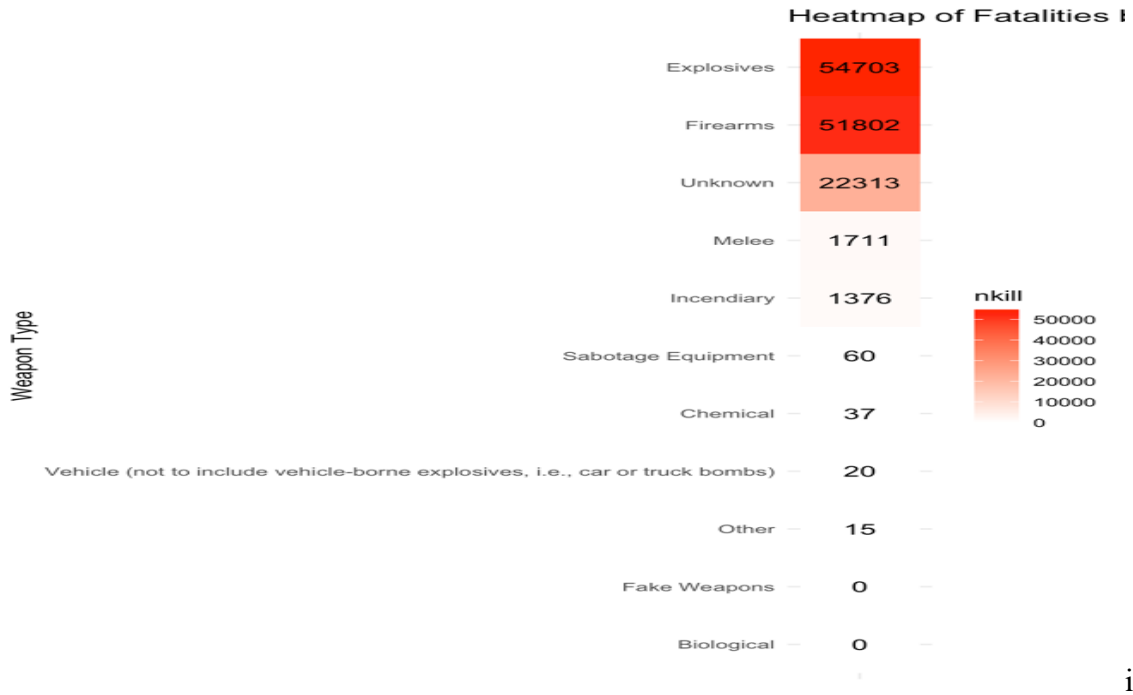
This heatmap illustrates the number of fatalities caused by different types of weapons used in terrorism attacks. The darker the color, the higher the number of fatalities associated with that weapon type.

• **Explosives (54,703 deaths)** and **Firearms (51,802 deaths)** are responsible for the highest number of fatalities, making them the most lethal weapon types used in attacks.

• **Unknown weapons (22,313 deaths)** also account for a significant number of fatalities, highlighting data gaps in attack reporting.

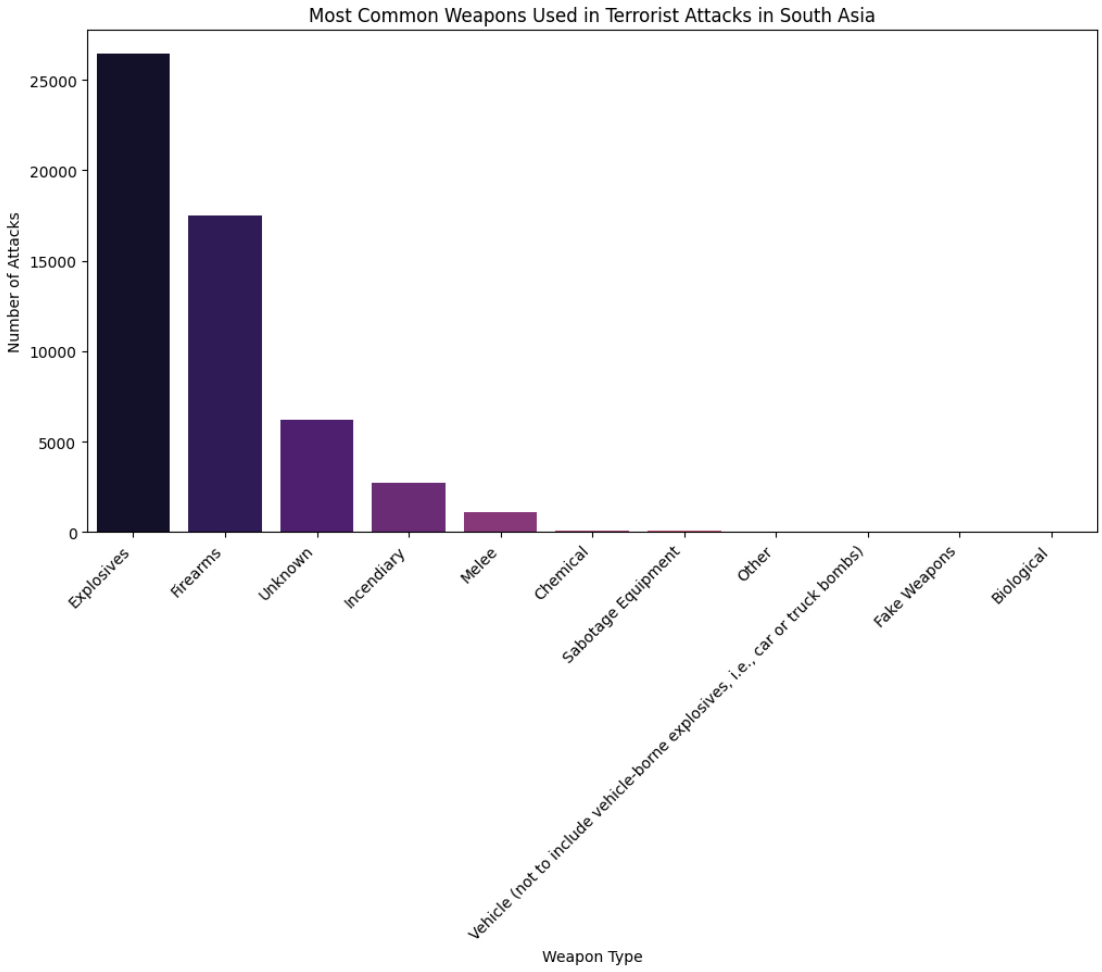
• **Incendiary (1,376 deaths)** and **Melee weapons (1,711 deaths)** have a relatively lower fatality count.

• **Biological and Fake Weapons** resulted in no fatalities, while **Chemical weapons (37 deaths)** and **Sabotage Equipment (60 deaths)** caused minimal fatalities.



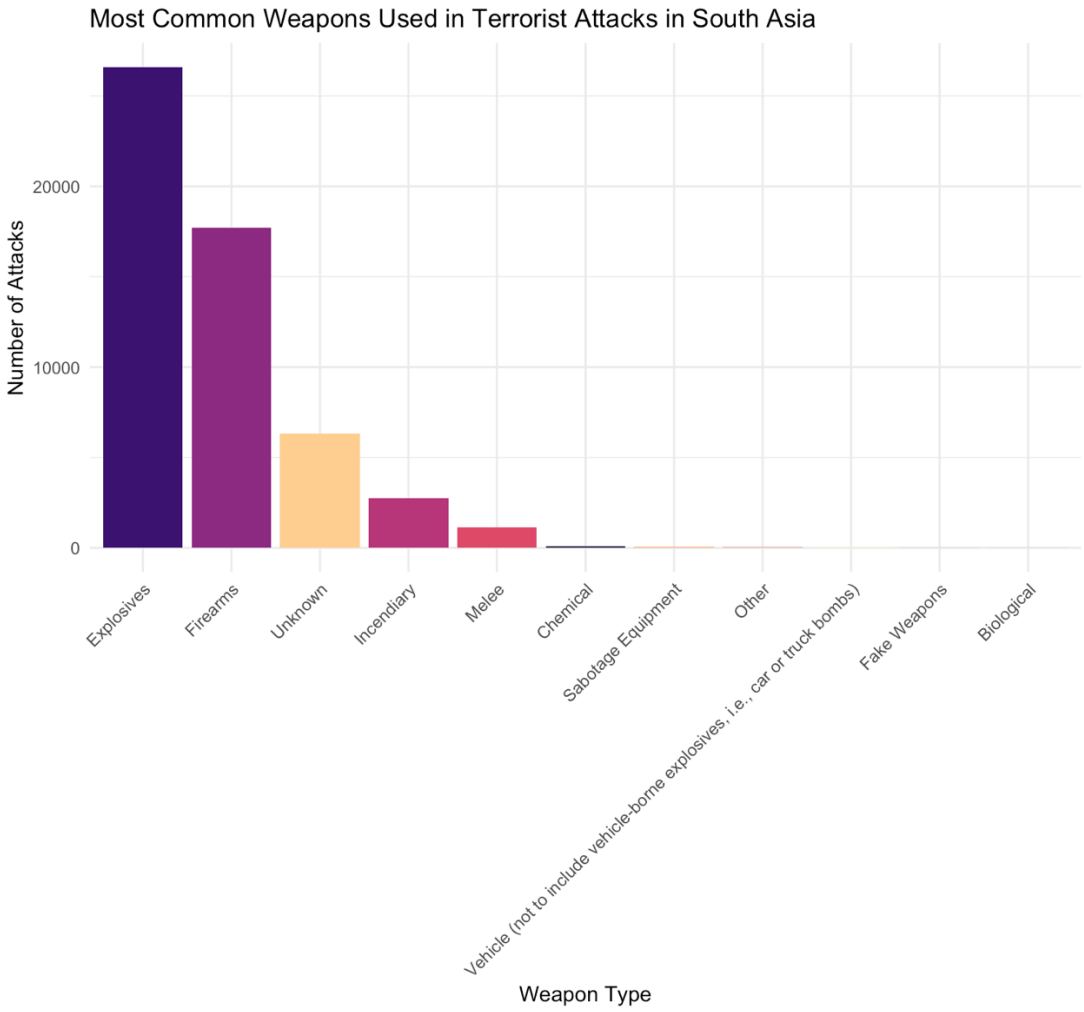

img provided by R

## MOST COMMON WEAPONS USED BY TERRORIST GROUPS




Img by python

Data Visualization HomeWork1.



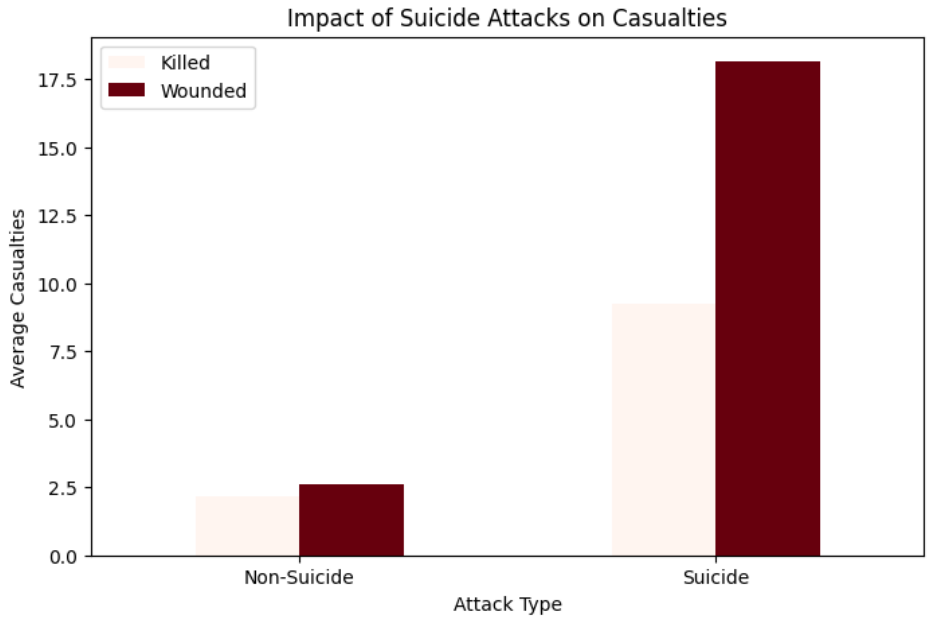

Img by R

**INFERENCE:**

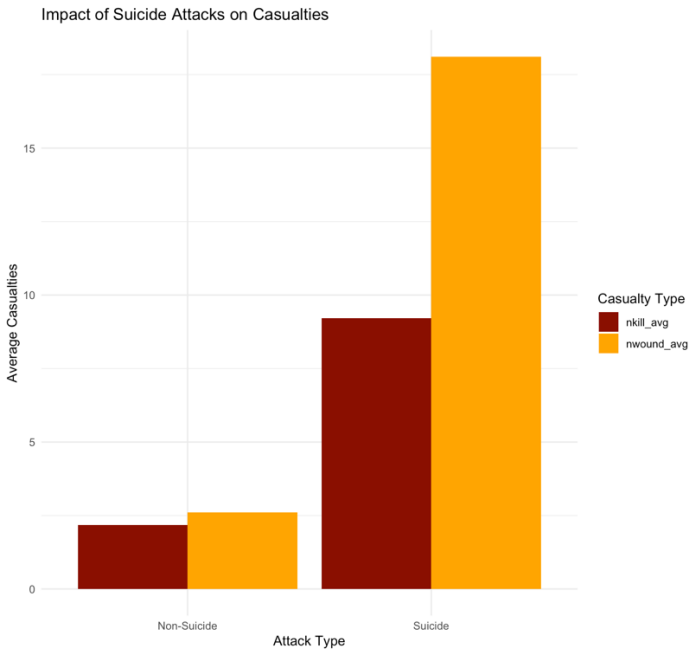
Explosives and firearms are the most commonly used weapons in terrorist attacks in South Asia, with explosives leading by a large margin. Unknown weapon types and incendiary devices also contribute significantly, while chemical, biological, and fake weapons are rarely used. This suggests a heavy reliance on high-impact, destructive tactics rather than unconventional or low-tech methods.

Data Visualization HomeWork1.

# IMPACT OF SUICIDE ATTACKS ON AVERAGE CASUALITIES




img by python

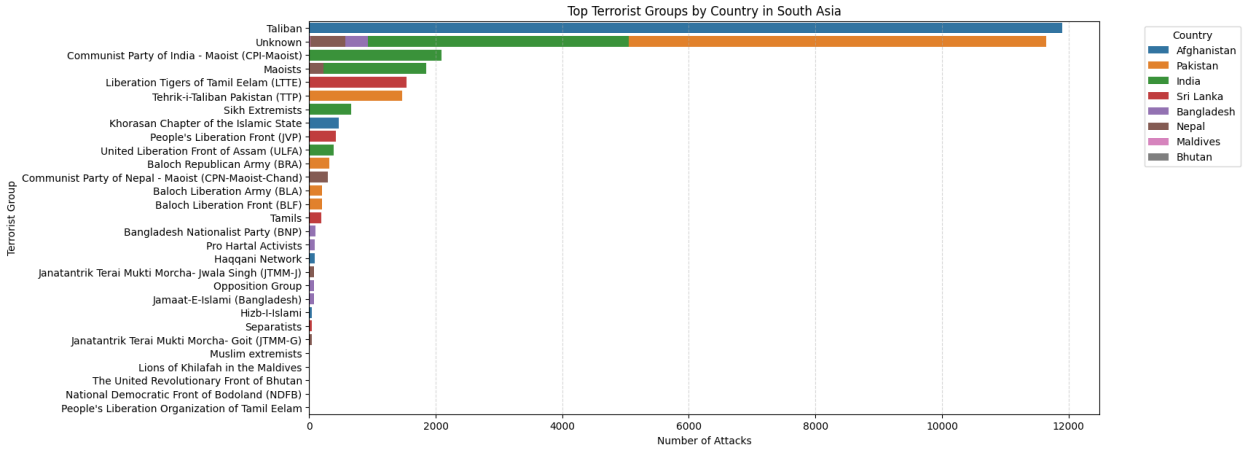



img by R

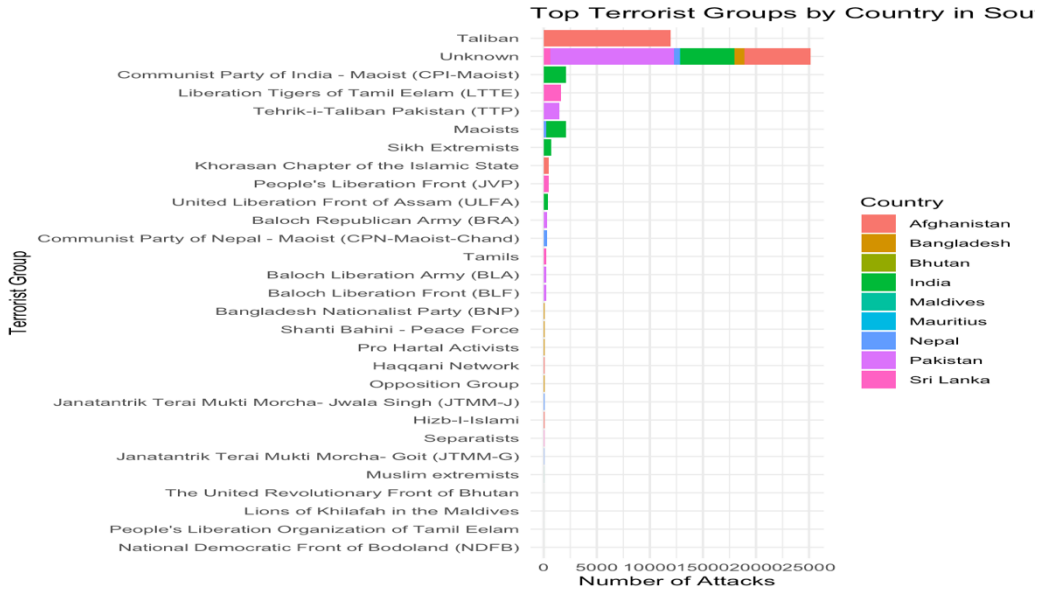
**INFERENCE:**

Suicide attacks cause significantly more casualties compared to non-suicide attacks. On average, both fatalities and injuries are much higher in suicide attacks, highlighting their devastating impact. Non-suicide attacks result in fewer casualties, making them less deadly in comparison. This suggests that suicide attacks are strategically used for maximum destruction and loss of life.

TOP TERRORIST GROUPS ACTIVE IN COUNTRIES:




Img by python

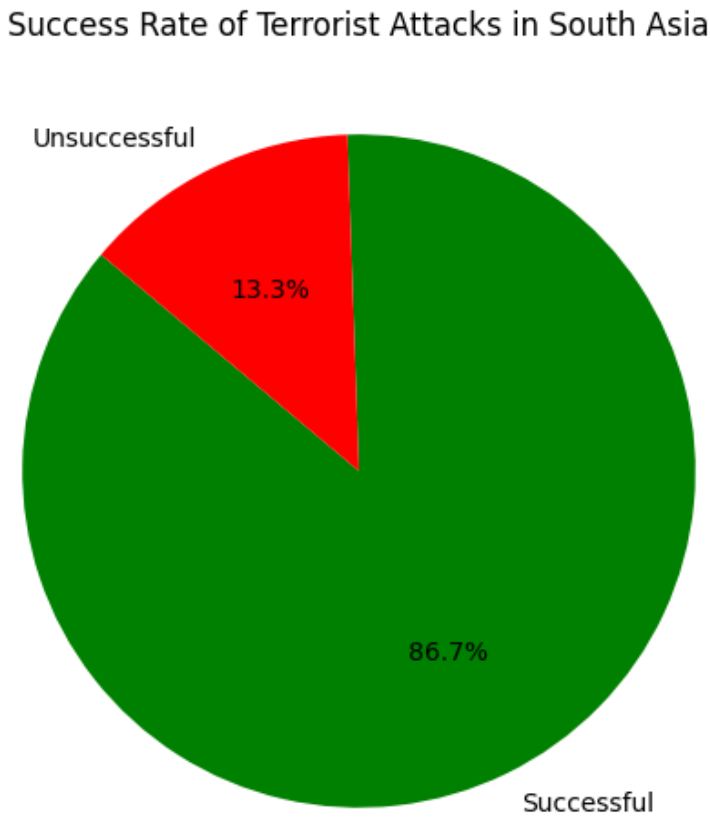



Img by R

**INFERENCE:**

The Taliban is the most active terrorist group in South Asia, with Afghanistan seeing the highest number of attacks. Pakistan has a significant number of attacks attributed to various groups, with many attacks recorded under unknown perpetrators. India has multiple active groups, including Maoists and separatist factions, contributing to its high number of incidents. Sri Lanka, Bangladesh, Nepal, and other South Asian nations also have groups responsible for attacks, but their numbers are comparatively lower.

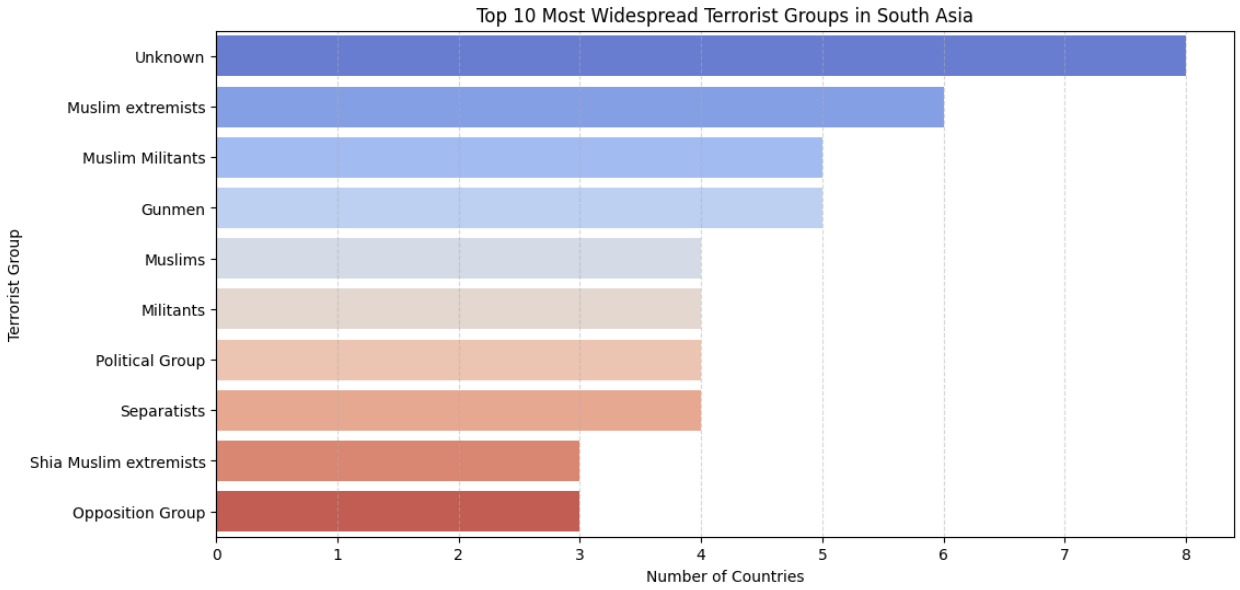
**SUCCESS RATE OF TERRORIST (goal accomplished)**




**INFERENCE:**

The majority of terrorist attacks in South Asia are successful, with an 86.7% success rate. Only a small fraction, around 13.3%, are classified as unsuccessful. This indicates that most planned attacks achieve their intended impact. The high success rate suggests a strong operational capability of terrorist groups in the region.

## INTERNATIONAL ACTIVITIES OF TERRORIST GROUPS

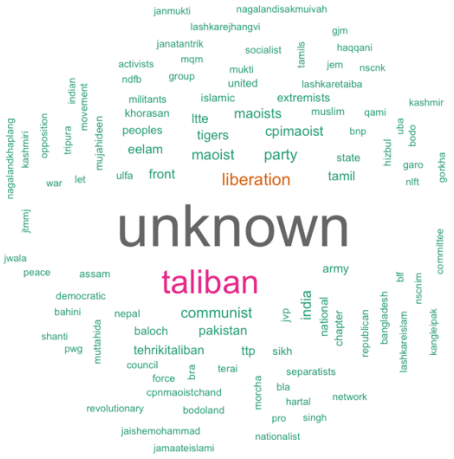



Img by python

## INFERENCE:

The most widespread terrorist groups in South Asia operate across multiple countries, with some groups being active in up to eight nations. Many of these groups are broadly labeled, such as “Muslim extremists” and “Separatists,” indicating a diverse range of actors. The presence of “Unknown” as the most widespread category suggests gaps in intelligence or unclaimed attacks. Overall, terrorism in the region is not confined to a single ideology or group but spans various movements and factions.

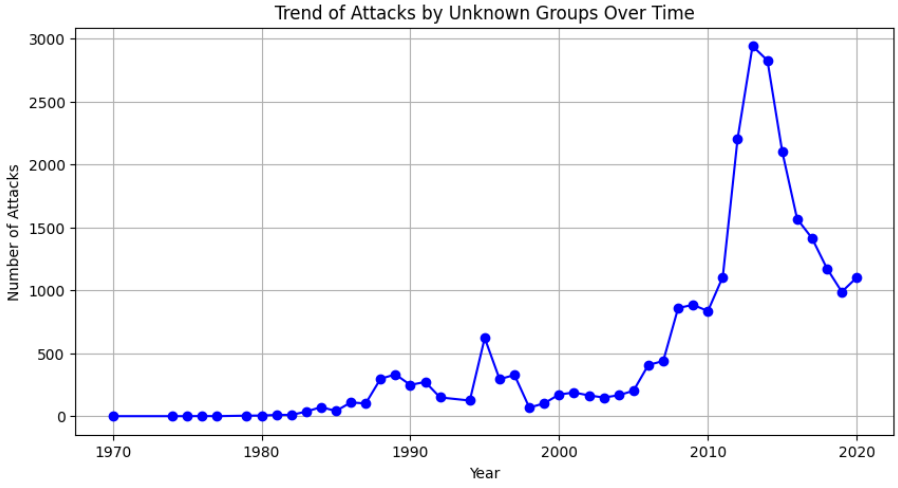

Worldcloud by python




Word cloud by R

So we infer that most of the Terrorism has not been accepted by any groups. Or it can be mix of the groups.

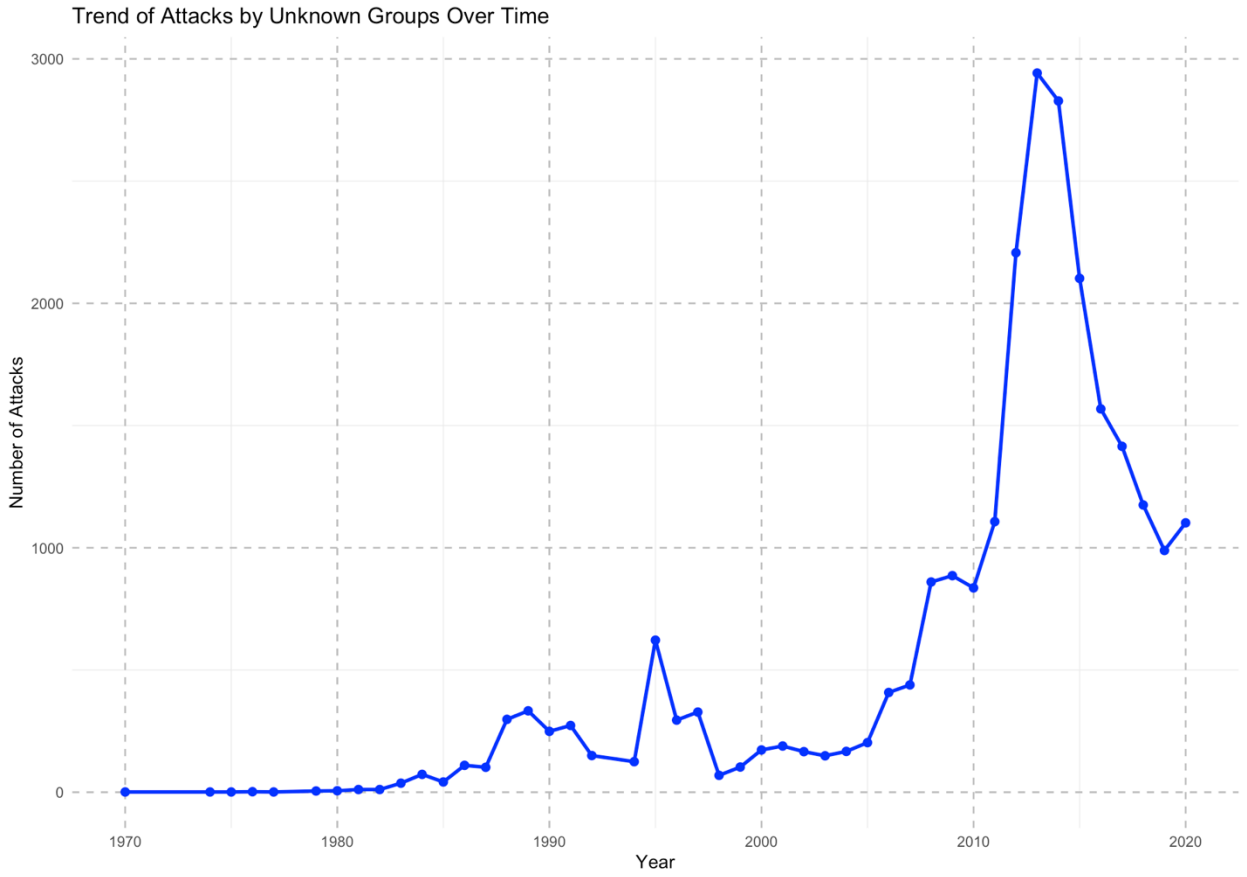
## **TRENDS OF ATTACKS OVER TIME BY THE UNKOWN:**




img by python

Data Visualization HomeWork1.

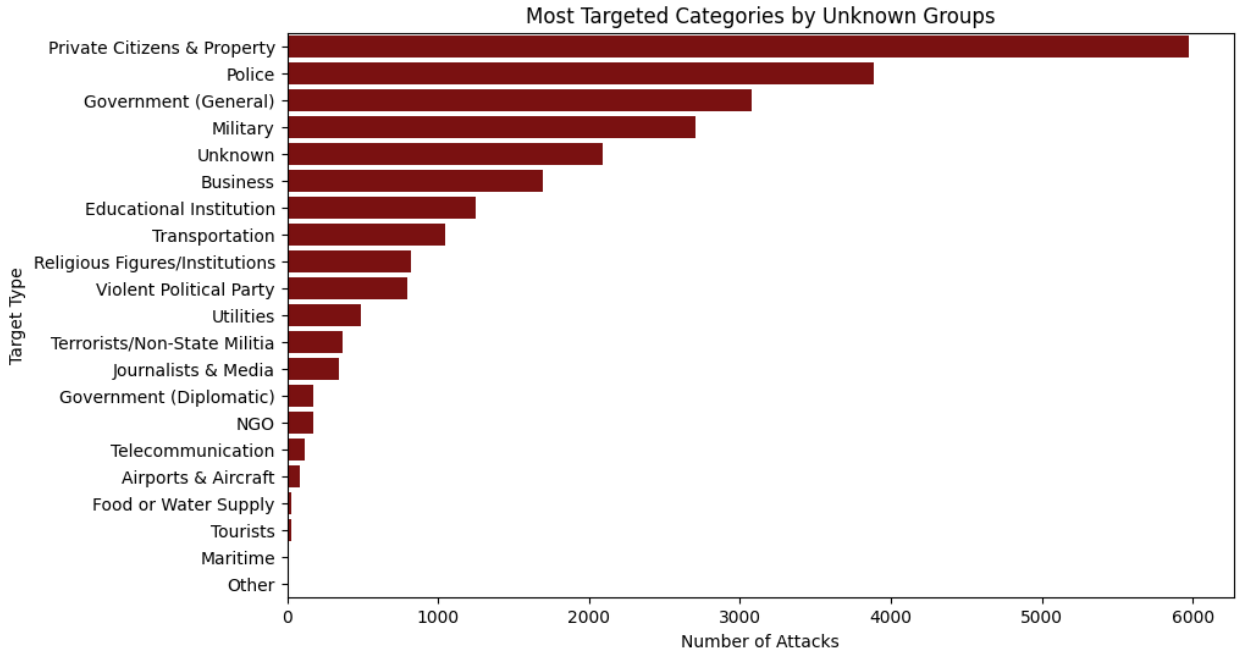
Img by R




**INFERENCE**

The trend of attacks by unknown groups shows a gradual increase from the 1970s to the early 2000s, followed by a sharp rise around 2010. The number of attacks peaked between 2012 and 2015 before declining significantly. This pattern suggests periods of heightened instability, potentially linked to regional conflicts or political turmoil, with a noticeable decrease in recent years.

## COUNTRIES TARGETED BY THE UNKOWN GROUP:

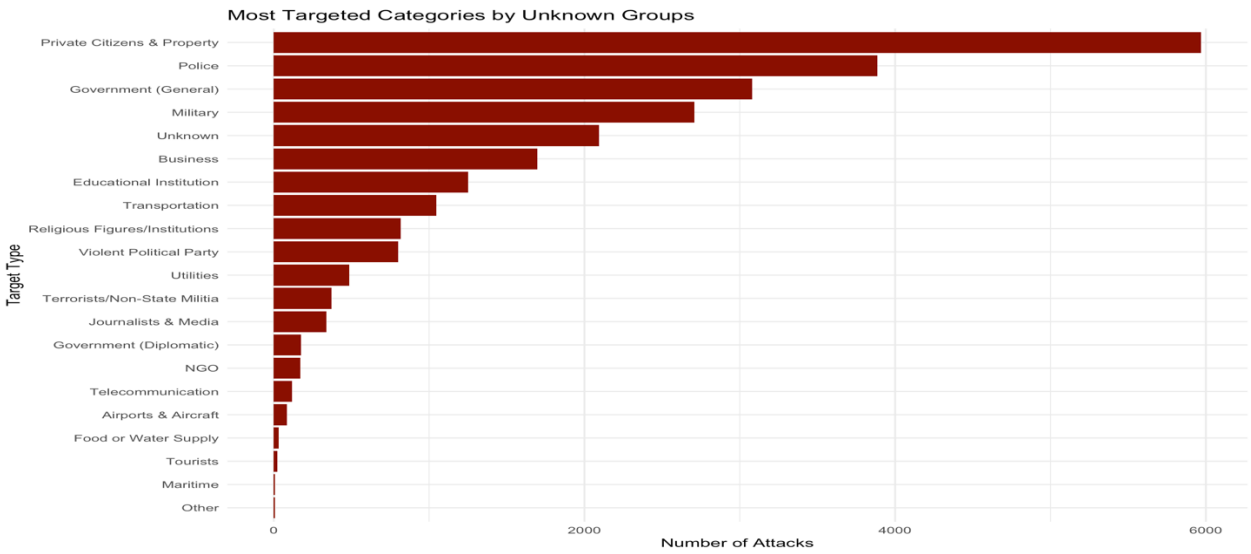



IMG BY PYTHON

## INFERENCE:

The most targeted categories by unknown groups indicate that **Private Citizens & Property** are the primary victims, followed by **Police, Government (General), and Military**. This suggests that these groups tend to attack public institutions and security forces, as well as civilians. Other notable targets include **Businesses, Educational Institutions, and Transportation**, highlighting disruptions to economic and social infrastructure. The lower frequency of attacks on **NGOs, Journalists, and Diplomats** suggests a relatively lower direct focus on these entities compared to state and civilian targets.

Data Visualization HomeWork1.




Img by R

**CONCLUSION:**

“Based on the analysis of the terrorism attack dataset, it appears that certain groups, including the unknown group, exhibit distinct patterns in terms of attack frequency, geographical distribution, and target selection. The unknown group remains a critical factor in understanding broader global security concerns, as their activities are often marked by unpredictable patterns. Further research is necessary to identify underlying motivations and gain insights into their structure. Identifying such groups will help in improving prevention strategies and addressing the root causes of terrorism.”