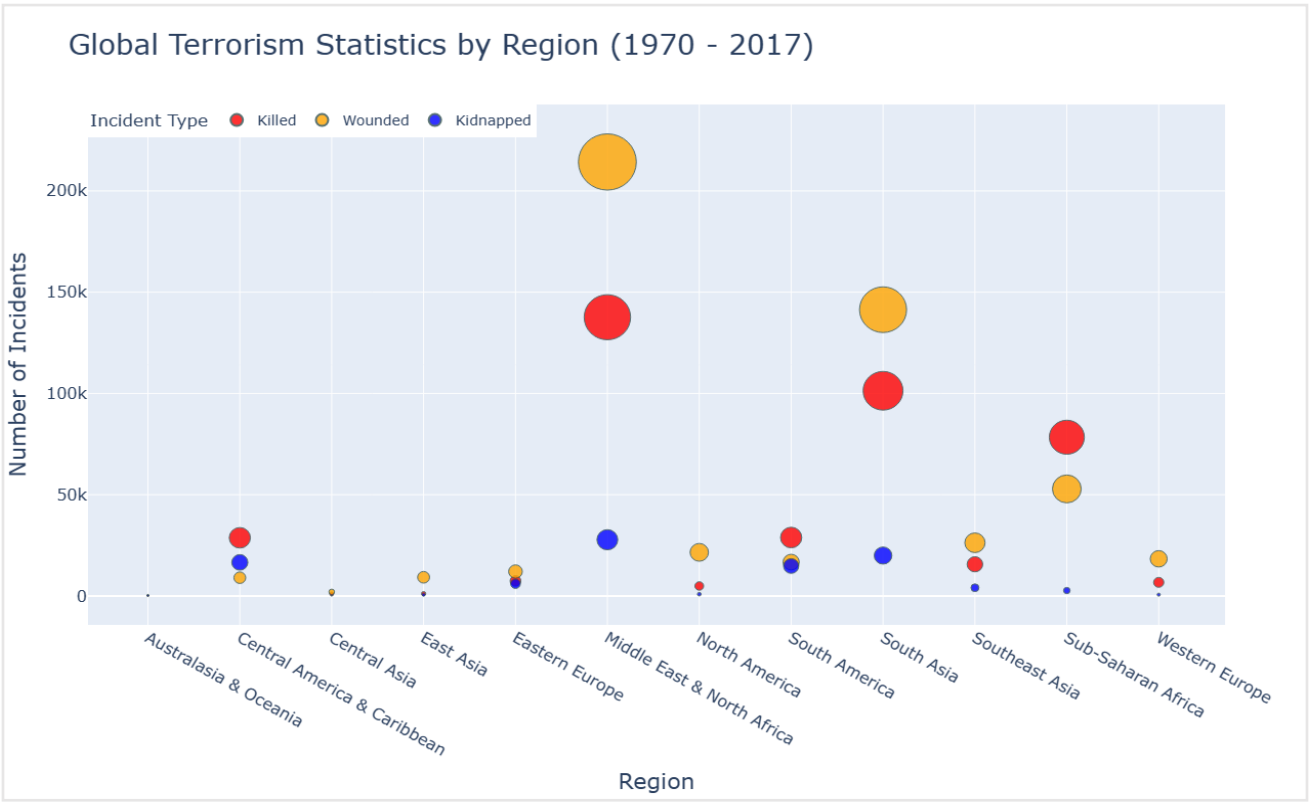
2151463, 2340638, 2337862
Phase 3 - Design

Goal: To visualize, analyze and compare the impact of terrorism in different regions around the world, specifically focusing on the number of people wounded, killed, and kidnapped. An additional goal is to identify outliers.



Insight:

1. The Middle East & North Africa region has experienced the highest number of wounded, killed and kidnapped individuals due to terrorism incidents in all regions.
2. Australasia & Oceania has had the least number of incidents with minimum number of casualties and kidnappings.
3. Kidnappings due to terrorism are relatively lower across all regions with Central America & Caribbean showing slightly higher numbers - the only region with more kidnapped than wounded individuals.

Outliers

1. Middle East & North Africa, South Asia, and Sub-Saharan Africa have seen significantly higher number of incidents (greater than fifty thousand) with much higher casualties (killed + wounded).
2. Australasia & Oceania and Central Asia have had a negligible number of incidents.

Point of interest:

1. Central Asia is surrounded by the two most volatile regions – South Asia and Middle East, yet the number of incidents, casualties, and kidnappings are surprisingly low.

Data Abstraction:

Dataset type: The Global Terrorism Dataset is a database of over 180,000 terrorist attacks, domestic and international incidents, worldwide from 1970 to 2017. START researchers at the University of Maryland keep it updated. The original Dataset is stored in a tabular form with 135 attributes and 181691 individual samples. The preprocessed dataset for this visualization has the following item and attributes:

- **Items** – Each row shows the total number of people injured, killed, or kidnapped by a terrorist attack in a region.
- **Attributes:**
    - Region – This is a categorical variable that shows the global area where a terrorist attack occurred. It has twelve categories.
    - Incident type – This is a categorical variable that indicates the kind of injury caused by a terrorist attack. It has three categories: Killed, Wounded, and Kidnapped.
    - Count – This is a quantitative variable; it represents the total number of individuals affected.
    - Number of incidents – This is a quantitative variable that shows the total number of incidents that occurred in a region.

Task Abstraction:

- **Marks:**
    - Point – Represents a single data value on the coordinate system. In our visualization, the point marks show the number of incidents of different incident types in each region.
- **Channels:**
    - Color hue channel – Represents the type of incident. Red for Killed, yellow for wounded and blue for kidnaped.
    - Vertical spatial position channel – Shows the number of incidents.
    - Horizontal spatial position Channel – Used for region attribute where the incident took place.
    - Size (area) channel – Size of the bubble or point mark represents the total number of people affected.
- **Users:** General public, researcher, journalist, and policy maker can make use of this graph to form policies, extract information, make or be aware of the state of terrorism in different regions. For example, a high number of attacks and casualties can be a call for military deployment, dispatching more ambulances or warning the public to take more precautions.
- **Actions:**
    - High-level – Present.
    - Mid-level – Lookup.
    - Low-level – Compare.

  This scatter plot allows the user to look up the number of people killed, wounded, or kidnapped in each region due to terrorist attacks. It also enables the user to compare the different regions and incident types based on the size and color of the dots.
- **Targets:** Outliers and Features – the target of interest to a user are features of the bubble, for example area, color, and vertical height of the bubble. Users can also identify outliers based on points that do not fit in with other points.

Encoding:

1. **Color coding:** Different colors represent different types of incidents (red for killed, yellow for wounded and blue for kidnaped). This makes it easy to distinguish between the types of incidents.
2. **Size of bubbles:** The size of the bubbles corresponds to the number of incidents, making it intuitive to understand the magnitude without reading specific numbers.
3. **X-Axis categorization:** The regions are clearly labeled on the x-axis, allowing users to quickly identify geographical trends in terrorism statistics.
4. **Y-Axis quantization:** The height or vertical position of the bubble represents the number of incidents that have occurred. It immediately shows the impact on a region in terms of the number of lives affected.

Additional Data Source:

URL: Global Terrorism Database (kaggle.com) – The data is stored in a single CSV file and it is available to be downloaded directly from Kaggle.

Tool used: This bubble visualization was created using Plotly Python graphing library on Jupyter Notebook.

Steps to recreate: Please follow the instructions in the readme.md file to recreate the visualization.