The United Nations estimates nearly 70% of the world’s population will be living in urban areas by 2050. Living in cities has obvious benefits such as public transportation infrastructure, attractions and shopping, and access to medical services. On the other hand, disadvantages include increased costs of living, noise and light pollution, and limited space. Data science provides us with a multitude of methods to quantify the pros and cons of urban dwelling. The process allows us make informed decisions using insights gleaned from analysis.

For this week’s article, I sought to determine whether cities offer increased safety when compared to rural areas. If this were to be the case, I next wanted to know which countries offered the safest cities. I arrived at this concept after completing a thorough comparison of different datasets. Instead of working backward from the narrative to the data, I decided to build a story around data. The data source used in this article contains more than 180,000 records. This provides more than ample information to create a narrative and derive meaningful insights.

The first step in attempting to answer the question of safety by city and country requires choosing the right indicators. This decision was made to be made by the data I had access to. [Kaggle](http://www.kaggle.com) is well known in the data science realm for maintaining a plethora of datasets ranging from wine reviews to baseball statistics. I selected the [Global Terrorism Database](https://www.kaggle.com/START-UMD/gtd), a comprehensive dataset maintained by the University of Melbourne’s National Consortium for the Study of Terrorism. The database contains detailed information about causalities, motivation, perpetrators and geographical information.

The next step involved reading the Excel sheet with Python in a Jupyter Notebook. This particular dataset includes different levels of geographic indicators. To establish which countries had the highest rates of terrorist attacks, I decided to group by region. This was done using a Python package called [Altair](https://altair-viz.github.io/). This visualization library offers a robust feature set and impressive visuals.

**<Code and image of nest donut>**

The image above contains countries belonging to the three regions with the highest amount of terror attacks. Notably, France, the United Kingdom, and the United States, among others are excluded due to their respective regions not having experienced as many cumulative incidents. Instead of using raw numbers, this visual gives an idea of where attacks most often occur and identify the biggest regional contributors.

After establishing an foundation of roughly where attacks occur, the next step involves exploring the data to utilize another indicator. The pie chart in the following image depicts the proportionality of different types of attacks. While the data source has different levels of details for attack types, only the highest-level categorization was used for simplicity. Matplotlib, another Python library, includes a technique known as “explosion” which allows for wedges to jut out.

**<Code and image of types of attacks>**

When working with particularly slim slices from the pie chart, this technique allows for an easy visual intuition. Bar charts offer another option for displaying correlated groups, however, I felt as though a pie chart better exhibits proportionality.

While we have established a cursory understanding of countries comprising the lion’s share of global attacks, only a map can provide definitive answers for this case. The Global Terrorism Database stood out from other datasets because it contained longitude and latitude coordinates for every record. These coordinates are easily the most accurate geo-mapping option. Other datasets contained records with varying degrees of coordinate information. On a global-mapping scale, ambiguity, incomplete names, and variation in spelling greatly constrain the ability to use other indicators such as cities and towns. Programs such as Tableau are adept at determining locations using the these indicators, however, on a scale of more than 180,000 records, other approaches must be taken. In this particular case, there were still approximately 30,000 rows which the program could not interpret.

**Tableau Terrorist Map**

The aforementioned longitude and latitude details allow for pinpoint accuracy in the map above. The amount of casualties is used to determine the size of each incident. Surprisingly, a significant number of the attacks had no causalities. This map reinforces the findings of the first chart; Afghanistan, Iraq, Pakistan, and India all experienced significant levels of attacks. The first graph, however, fails to depict the other countries like the Philippines, Germany, and many more in Central American who also ranked highly on the index.