**Date:**     November 5, 2017

**To:**        Professor Samuel Johnson

**From:**    Jianhui Lu, Weiran Yuan

**Subject:** Proposal of Terrorism and College Safety Application


Safety is one of the most important factors of choosing a college that we are going to spend several years life, terrorism would be one of the greatest threatens to United States and the world, so our group is connecting these two factors together to provide a reference to the safety of a specific college. We will provide colleges information and terrorist attacks relevant to the college. The target users of the application would be the students going to university or their parents.


## Datasets information

The datasets that we need in the development of the application are terrorism data and colleges data.

The terrorism dataset we found is global scope, it records more than 170,000 terrorist attacks reported on news and complete information about the attacks, including year, date, country, province or state, city, coordinates, attack type, target and weapon, etc. of the terrorist attack. The source of the terrorism dataset is Global Terrorism Database, copyright belongs to University of Maryland. It's an open-source database, so the data is open to the public and we are allowed to download the dataset, but we should cite the database each time we use it. And we cannot reconstruct the data or propagate it on the internet.

Source of the terrorism dataset: http://www.start.umd.edu/gtd/

License of the terrorism dataset: http://www.start.umd.edu/gtd/terms-of-use/

The colleges data is domestic scope in United States, it contains basic college information such as name, zip code, city, Accred Agency, official websites and coordinates. The source of the colleges data is federal government data under education catalog, it's a public database and permit us to use all of its data for any purpose.

Source of the colleges dataset: https://catalog.data.gov/dataset/college-scorecard

License of the terrorism dataset: https://creativecommons.org/publicdomain/zero/1.0/legalcode

These two datasets have common attributes of city and coordinate, sometimes the terrorist attack targeted at the location of a college. We plan to join the two datasets by their common attribute city, a terrorist attack should happen in a specific city and if it happened recent enough, the safety of the college should be considered carefully from our perspective, since the students of the college should be most active in the city. But this should be a reference for the users, so we also included other information of the college for them to have a basic conception.

On the other hand, if the user expects to study more about the terrorist attacks, we also provide a choice to show more information of the attacks. The schema of the terrorism data should be long enough for users to learn about the attacks, but a detailed story is not included in it, the users need to google it themselves.

**Programming languages**

We will use python and PHP as the main languages to develop the application and connect the database. Part of the data will be stored in MongoDB, a non-relational database, and the rest will be store in Postgres database. Our python or PHP program will be developed to connect both of the databases. The program will ask for key words from the users to construct a SQL query string or MongoDB query string to search the databases, the feedback will show on the interface, people can check things they need in the database.

We upload the two datasets on Githud, terrorism dataset named globalterrorismdb_0617dist.xlsx and college dataset named CollegeData.csv. You can also check it on the URL:

https://github.com/ianLo93/Database-system-project

We also attached these two files on LMS memo submission.