**Project I**

**Missing Migrants Project Proposal**

**Shreya Prabhu(shreyapr) & David Lee(dlee3)**

**10/01/2019**

**Motivation:**

We chose the missing migrants project given the current global political climate and increasing tension. While the crisis that form the basis for mass migrations today, are themselves devastating, the missing migrant framework provides a new perspective on how we look at the crisis. We believe this dataset to be interesting to find some patterns with. While this problem is mainly due to personal interest, we believe that the findings could have value for organizations and governments to investigate more into the causes for people going missing and work towards that. We are hopeful that these findings can also help reduce the recurrence of these tragic incidents.

**Related Work:**

Our team members have been interested in the refugee crisis for some time now and has been following the work of the International Organization for Migration. They did similar work during the Nepal Earthquakes of 2015, where they applied core data analysis to refugee settlements and used those patterns to aid resource allocations.

**Data:**

We sourced the data from Kaggle.com. Link: <https://www.kaggle.com/snocco/missing-migrants-project>

The data provides detailed information on the location and region of incident and also the demographics. The data tracks the deaths of missing immigrants as well as asylum seekers who have gone missing along their migration routes. It is essential for us to remain cognizant of the fact that these are minimal estimates and we are aware that not all deaths, especially in crisis situations, are well recorded. While this is currently the only source and dataset we have, we hope to continue our search for more information that can guide our project.

The dataset has 20 columns and over 5000 rows.

From our initial examination, the data seems to be fairly clean. However, we might choose to wrangle the rows and columns a bit, in order to make the analysis more efficient and convenient.

The [twitter](https://twitter.com/missingmigrants?lang=en) page for the movement maintains a steady stream of visualizations supported by this data. The [IOM](https://gmdac.iom.int/map-tracking-migrant-deaths-and-disappearances) themselves have provided a few visualizations in this regard.

**Possible Questions:**

* What are the most common causes of death among refugee migrants?
  + We plan to group the cases by the cause of death and rank them in order by total number of people
* Does any location have particularly higher or lower occurrences of death vs missing? If so, what could be the potential causes for that?
* The migration routes that have major incidents of death?
* The migration routes that have major incidents of missing people?
  + For the three questions above, we plan to investigate cases that have missing migrants and analyze if there is any difference between those and other cases, especially in regards to region and reliability
* Demographics of the people that have gone missing: Male/Female/Children
  + We believe this can be achieved using the Male/female and children columns to correctly identify the number
* Region of incidents and the crisis they are moving away from
  + We plan to map the trajectory from the region affected to the lat/long location

**Possible Findings and Implications:**

Overall, we expect (hope) for there to be a small number of people missing. However, it is absolutely essential to hold accountability for the missing people. In that regard, we expect to find more adults than children that will be missing. – realizing that children would be accompanied by parents in most cases.

At this point, we expect the Mediterranean region to have the greatest number of missing migrants, due to the massiveness of the Syrian crisis.

We also expect a fair amount of ‘unknown’ causes of death, realizing the situations in which these data points are recorded.