So from Dr. Dundar’s perspective, a K-means clustering algorithm is the only real tool I will need built in that I will need before doing analysis on crime patterns.

“From my perspective a simple Kmeans clustering algorithm would do the job. However, this project is interesting mainly because of the potentially interesting patterns you could discover.”

Furthermore …

1. Cluster all crime vectors across all years and cities to identify crime patterns and manually analyze them to see if they make sense. Which crime pattern is least desirable, which one is relatively more tolerable? Do the crime patterns assigned to cities we know make sense? If not, then can we adjust clustering to more sensible assignments?

Essentially, do the K-means clustering algorithm, then look at years and cities lumped together in similar clusters (15 clusters), then look at those clustered cities over time, do the patterns make sense? Then, which of these clusters is the least favorable (worst in terms of caliber), and which are relatively better (not as awful)? Examine Indianapolis to see if this type of crime pattern makes sense for Indianapolis. If not, we should adjust the K-means algorithm.

1. Once we make sure the number and the type of crime patterns make sense we can investigate temporal crime trends across specific cities to identify interesting patterns.
   1. For example there are big cities with 100K+ populations that has the same crime patterns as they did 20 years ago when their populations was less than 10K. These are cities that developed healthily. On the other hand there are cities which switched to worse crime patterns as they develop. We can identify these cities if we look into temporal pattern assignments. What method can we use there?
   2. There are big cities (population flat) which changed for better or worse after certain years. Can we identify those cities? Can we look into some external factors that may cause this change?

Essentially, if the patterns from (1) make sense, then we want to begin looking into a specific city for crime trends to identify interesting patterns. (a) One interesting pattern we could look for is **whether a city has developed healthily over time**. An example of this is a city with *a* populationover 100,000+ that has the same crime patterns as they did 20 years ago when their population was less than 10,000 we would consider having ***healthy growth***. **An opposite example would be a city that switch to worse crime patterns as the population grew**, I am not yet exactly sure what method we can detect those cities, but I have a very good resource that may nudge me in the right direction (<https://www.cs.cornell.edu/people/tj/publications/shaparenko_etal_05a.pdf> ). Another good next google search would be, “*plotting clusters over time* *in matlab*” (<https://www.mathworks.com/help/stats/clustering.evaluation.silhouetteevaluation-class.html> ). **And finally, the last possibility is those cities that are 100,000+ but they did not grow or decay but remained flat, which changed for better or worse over time**. I am not sure yet how to identify those cities who were flat in terms of population, but got better or worse over time.

