**Revised Proposal**

Using the UNHCR [Refugees R package](https://www.unhcr.org/refugee-statistics/insights/explainers/refugees-r-package.html)

Dataset: asylum decisions

**Research Question: Which factors affect asylum decisions for refugees?**

1. Using multiple linear regression, find:
   1. Which factors are statistically significant (which factors affect asylum acceptance the most)
   2. Build a machine learning model to compute likelihood of acceptance
2. Dependent variable = asylum application acceptance rate
   1. The continuous variable Y = (dec\_recognized + decision\_other)/dec\_total
      1. dec\_recognized: The number of asylum applicants recognised as refugees under UNHCR’s mandate.
      2. dec\_other: The number of asylum applicants recognised as refugees under other forms of complementary protection.
      3. dec\_total: The total number of asylum applicants for which a decision on their asylum claim has been made in the given reference period
3. Independant variables:
   1. year. The year (e.g. 2020).
   2. app\_type
   3. coo\_name: Country of origin name
   4. coa\_name: Country of asylum name
   5. procedure\_type: The procedure type describes the authority with whom the asylum claim was lodged. It can be one of: G – Government, J - Joint, U – UNHCR.
   6. dec\_level: The decision level provides more detail of the stage of the procedure. Common values are NA (New applications), RA (repeat/reopened applications) and AR (administrative review).
   7. dec\_pc: The data type describes whether the data provided reflects cases or persons. The most common situation is for the data on asylum applications lodged to reflect persons, but in certain situations (e.g USA), the data provided by governments is at the level of cases. A case can contain information on one or more persons.