Hierarchical Bayesian Regression modelling
Crime Prediction in Uttar Pradesh (District wise)

1. **Data Modelling**:

   (a) *Data source*
   The **GDP** (Gross Domestic Product) data was collected from Open Government Data Platform India based on constant prices from the year 2004 to 2011.
   Number of Males in the age group (15-24) for the year 2011 was collected from UttarPradeshStat.com and the same data projected for the year 2016 was collected from Health Management Information System under Ministry of Health & Family Welfare, Government of India. We will also be calculating **Gender ratio** using above data for the year 2011 and 2016. **Unemployment** data is collected from UttarPradeshStat.com for the year 2011. We can make use of unemployment rate of Uttar Pradesh to project unemployment data for the years 2013, 2014, 2015.
   Available data is of district level granularity for the state of Uttar Pradesh, unless mentioned explicitly.

   Upon realizing that total crime count likely depends upon recent activities and the time of the year, autoregressive terms for total crime count are added. Trailing (a month and a year) crime counts are also considered as predictors.

   A clustering based on per capita GDP, Unemployment, and Population, will be used to group districts with similar patterns of total crimes committed, and we will try to relate districts with similar geographic context.