

## CSCI 4380 Database Systems Project Memo

### Team members:

Jizhou Huang  
Jihui Nie  
Yitao Yang

### Datasets:

1. Unemployment statistics in New York State beginning from 1976  
(76,200 rows  $\times$  7 columns, provided by New York State Department of Labor)  
<https://data.ny.gov/Economic-Development/Local-Area-Unemployment-Statistics-Beginning-1976/5hyu-bdh8>
2. Crime statistics in New York State beginning from 1990  
(19,300 rows  $\times$  15 columns, provided by New York State Division of Criminal Justice Services)  
<https://data.ny.gov/Public-Safety/Index-Crimes-by-County-and-Agency-Beginning-1990/ca8h-8gjq>

### Joining the datasets:

- The two datasets share two attributes in common: *county* and *year*. Though the former starts from 1976, the actual size of data between years 1976 and 1989 is very limited (~700 rows out of ~70,000 rows). Therefore we regard the former dataset as data since 1990.