**ETL Project Plan**

* Data
  + Chicago PubHealth Indicators 2015 csv
    - Transform
      * Select Columns
        + Community Area \* (PK)
        + Community Area Name
        + Unemployment
        + Dependency
        + No High School Diploma
      * Drop NULLS
      * Community Area – change type to INT
  + Chicago Crime 2018 csv
    - Transform
      * Select Columns
        + ID
        + Community Area \* (PK)
        + Primary Type
      * Drop NULLS
      * Drop rows where Community Area = 0
      * Community Area – change type to INT
      * Group by Community Area
  + Chicago Population Data 2017 csv
    - Columns
      * Community Area \* (PK)
      * Population
  + Merged Crime and Population
    - Merge on ‘Community Area’
    - Transform
      * Crime rate (Calculation)
        + Calculate crime rate

# crimes / population \* 1,000

* + - * Create variable – Drug Related Crime
        + Primary Type – Narcotics & Other Narcotics Violation
        + Combine these two offenses into new variable
* DB
  + Postgres
  + ERD Diagram
* Correlation
  + Crime rate and Unemployment
  + Unemployment and Education (No HS diploma)
  + Drug related crimes and Dependency