

# Data Modeling Project 1

#Proposal for Data Modeling

1. Jiwon Shin
2. Hoachong (Harry) Xia
3. Xueqing (Annie) Wu
4. Rafael Dávila Bugarín

## Data Set

The Data Source comes from Los Angeles Police Department (LAPD) and it is available in the following [link](#). In addition, migration data will be merged to answer one of the research questions, the data of migration comes from this [link](#) and contains County-to-County migration flows from 2016 to 2020.

## Data Set Description

This dataset represents recorded crime incidents in the City of Los Angeles dating from 2020. The source of this data is transcriptions from original crime reports, which are originally documented on paper. Therefore, it is important to acknowledge that there may exist some inaccuracies or errors within the dataset. In certain cases, location information is missing and indicated as (0°, 0°). To safeguard individual privacy, address details are truncated to the closest hundred block. The accuracy of this data is contingent on the accuracy of the information stored in the database. Please feel free to raise any questions or concerns in the comments section.

## Research questions

1. Does time frames really affect the crime rate in LA in 2022? (e.g. day, night, morning, afternoon, evening)
2. Does the crime rate in LA in 2021 affect the migration rate in 2022?

## Glimse of the data

```
Attaching core tidyverse packages          tidyverse 2.0.0
dplyr      1.1.2      purrr      1.0.2
forcats    1.0.0      readr      2.1.4
ggplot2    3.4.3      stringr    1.5.0
lubridate  1.9.2      tidyr      1.3.0
Conflicts          tidyverse_conflicts()
dplyr::filter()   masks stats::filter()
purrr::flatten() masks jsonlite::flatten()
dplyr::lag()      masks stats::lag()
Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

Rows: 1,000

Columns: 26

```
$ dr_no      <chr> "010304468", "190101086", "200110444", "191501505", "19...
$ date_rptd  <chr> "2020-01-08T00:00:00.000", "2020-01-02T00:00:00.000", "...
$ date_occ   <chr> "2020-01-08T00:00:00.000", "2020-01-01T00:00:00.000", "...
$ time_occ   <chr> "2230", "0330", "1200", "1730", "0415", "0030", "1315",...
$ area       <chr> "03", "01", "01", "15", "19", "01", "01", "01", "01", "...
$ area_name  <chr> "Southwest", "Central", "Central", "N Hollywood", "Miss...
$ rpt_dist_no <chr> "0377", "0163", "0155", "1543", "1998", "0163", "0161",...
$ part_1_2   <chr> "2", "2", "2", "2", "2", "1", "1", "2", "1", "1", "1", ...
$ crm_cd     <chr> "624", "624", "845", "745", "740", "121", "442", "946",...
$ crm_cd_desc <chr> "BATTERY - SIMPLE ASSAULT", "BATTERY - SIMPLE ASSAULT",...
$ mocodes    <chr> "0444 0913", "0416 1822 1414", "1501", "0329 1402", "03...
$ vict_age   <chr> "36", "25", "0", "76", "31", "25", "23", "0", "23", "0"...
$ vict_sex   <chr> "F", "M", "X", "F", "X", "F", "M", "X", "M", "X", "M", ...
$ vict_descent <chr> "B", "H", "X", "W", "X", "H", "H", "X", "B", "X", "A", ...
$ premis_cd  <chr> "501", "102", "726", "502", "409", "735", "404", "726",...
$ premis_desc <chr> "SINGLE FAMILY DWELLING", "SIDEWALK", "POLICE FACILITY"...
$ weapon_used_cd <chr> "400", "500", NA, NA, NA, "500", NA, NA, NA, NA, "306",...
$ weapon_desc <chr> "STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)", "UNKN...
$ status     <chr> "AO", "IC", "AA", "IC", "IC", "IC", "IC", "IC", "IC", "...
$ status_desc <chr> "Adult Other", "Invest Cont", "Adult Arrest", "Invest C...
$ crm_cd_1   <chr> "624", "624", "845", "745", "740", "121", "442", "946",...
$ location   <chr> "1100 W 39TH                                PL", "700 S HILL...
$ lat        <chr> "34.0141", "34.0459", "34.0448", "34.1685", "34.2198", ...
$ lon        <chr> "-118.2978", "-118.2545", "-118.2474", "-118.4019", "-1...
$ crm_cd_2   <chr> NA, NA, NA, "998", NA, "998", "998", "998", "998", NA, ...
$ cross_street <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "OLIVE", NA, NA...
```