Chicago Crime Data Analysis

Data Engineering Platforms for Analytics
Autumn 2022

Zoey Chen
Kanu Madhok

TABLE OF CONTENTS

EXECUTIVE SUMMARY

In the project, we analysed the data from “Chicago Crime dataset report.” This has the numbers and information of crimes for Chicago city that occurred from 2021 to analyze the trends of crimes.

There are several questions we would like to answer, such as which types of crimes are most frequently committed and which areas have the highest crime committed rate. We can also examine how different factors such as time play on crime. We can determine the impact different world events play on the rate and type of crime.

Data Tools

- **Python**
  - API calls
- **MySQL**
  - Loading, and analyzing data
- **Open Refine, Excel**

  Extracting, and clean data
- **Tableau**
  - Visualization for Charts
- **R**
  - Visualization of GIS data

Data Sources

- City of Chicago - Crime dataset: https://data.cityofchicago.org/resource/dwme-t96c.json
- IUCR code:
  - Chicago Police Department - Illinois Uniform Crime Reporting (IUCR) Codes
- Community area:
  - Boundaries - Community Areas (current) | City of Chicago | Data Portal
  - https://www.cmap.illinois.gov/documents/10180/126764/_Combined_AllCCAs.pdf/
- District: Boundaries - Police Districts (current) | City of Chicago
- Ward office: City of Chicago :: Ward Offices
- FBI code: Definition & Description of Crime Types

DATA COLLECTION

- API CALLS
- ASSIGNING TO DATAFRAME
- EXPORT CSV

To collect data from chicago's data portal we used an API call. We used api calls specifying which year of data we wanted to extract. This was done in a python environment and was loaded into a pandas dataframe. Through pandas functionality we were able to clean the data and change column names to fit our schema. After all changes were made we exported the dataframes into one large csv file.

```
import requests
import json
import pandas as pd


# response = requests.get("https://data.cityofchicago.org/resource/ijzp-q8t2.json").text
#  https://data.cityofchicago.org/resource/ijzp-q8t2.json?$where=year > 2001
response_2019 = requests.get("https://data.cityofchicago.org/resource/ijzp-q8t2.json?$where=year > 2018").text
```

DATA COLLECTION

In [5]:

```
1  import pandas as pd
2  from sodapy import Socrata
3
4  client = Socrata("data.cityofchicago.org", None)
5
6  results = client.get("dwme-t96c", limit=200008)
7
8  results_df = pd.DataFrame.from_records(results)
9  print(results_df.count())
10 results_df.to_csv('file_2021.csv')
11
```

```
WARNING:root:Requests made without an app_token will be subject to strict throttling limits.
```

```
date                    200008
location                195037
district                200008
block                   200008
y_coordinate            195037
latitude                195037
description             200008
location_description    199215
updated_on              200008
community_area          200008
iucr                    200008
x_coordinate            195037
ward                    199997
case_number             200008
year                    200008
domestic                200008
fbi_code                200008
longitude               195037
beat                    200008
primary_type            200008
arrest                  200008
id                      200008
dtype: int64
```

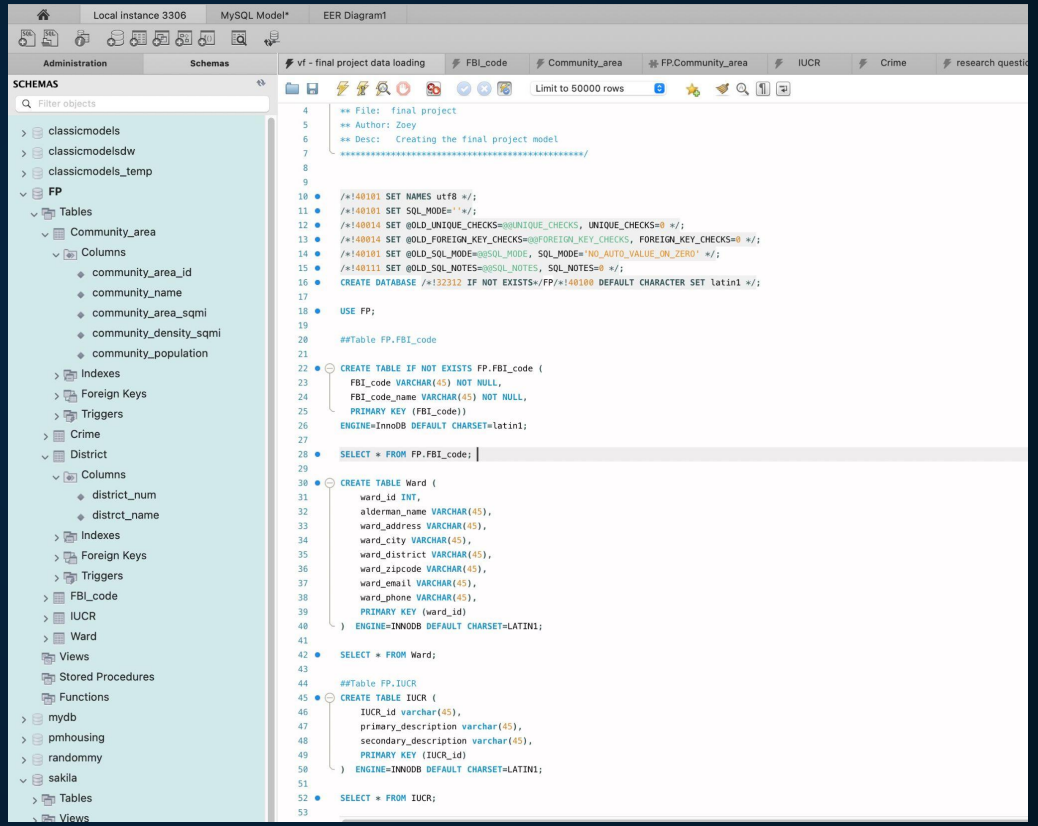
- We used endpoints to pull data through an api call
- This lead to a challenge as there were limits to how much data could be called
- Through research we were able to identify an easier and more efficient process of calling the api
- Looking through documentation an alternative approach was designed
- Api call exported data into a pandas dataframe which was wrote into a csv file

Data Extraction

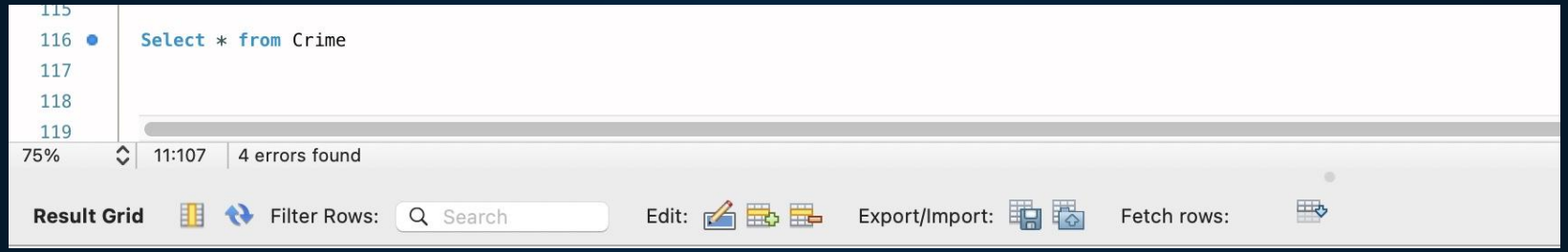
We mainly use Excel and OpenRefine to clean up our data, we deleted the columns we don't need in the large CSV file and make sure the data we have are crimes in 2021.

Data Loading

- SQL to create Schema
  - Create 6 tables to load data



Data Loading



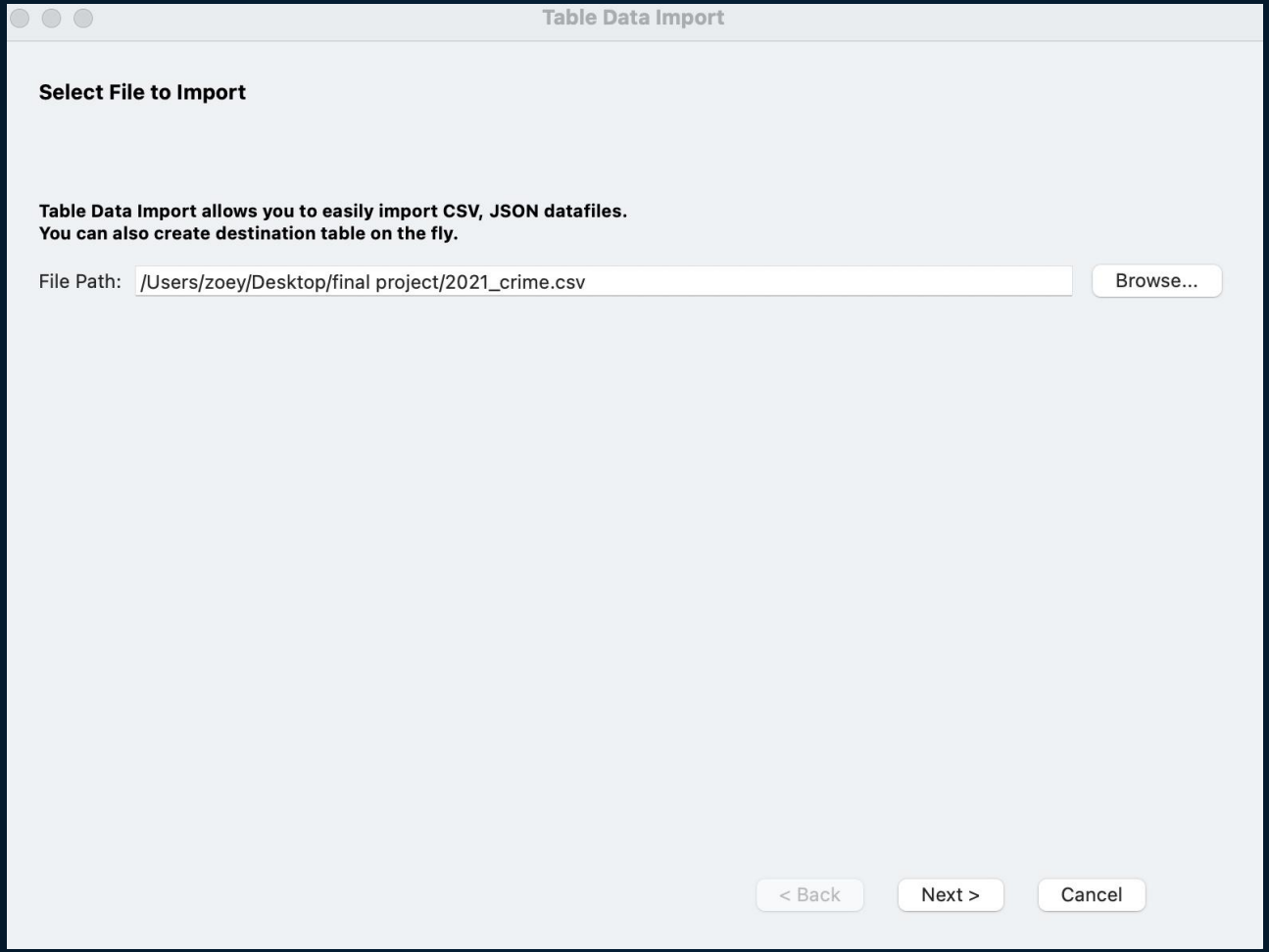


Table Data Import
Select File to Import
Table Data Import allows you to easily import CSV, JSON datafiles.
You can also create destination table on the fly.
File Path:
/Users/zoey/Desktop/final project/2021_crime.csv
Browse...
< Back
Next >
Cancel

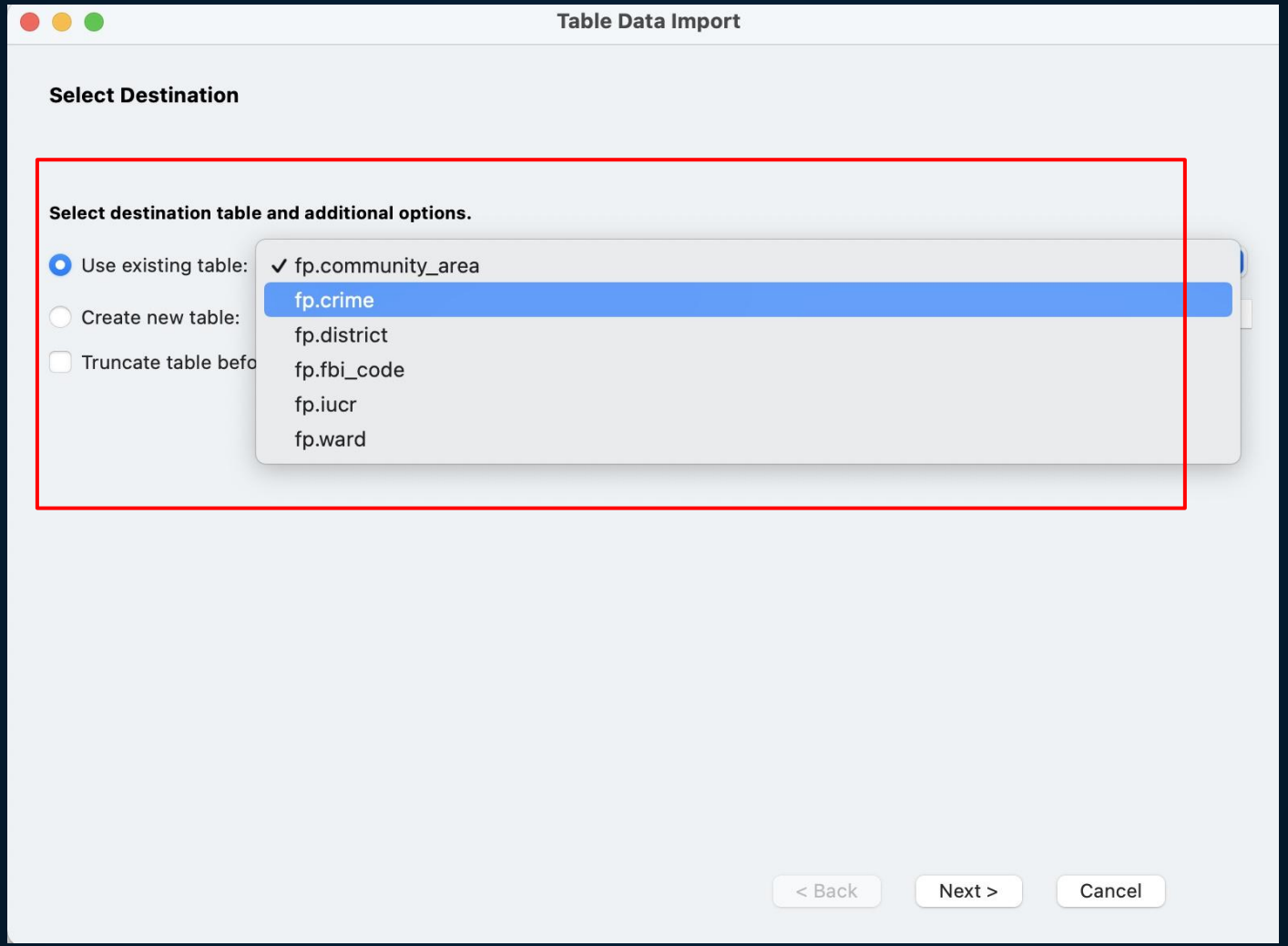


Table Data Import
Select Destination
Select destination table and additional options.
Use existing table:
Create new table:
Truncate table befo
fp.community_area
fp.crime
fp.district
fp.fbi_code
fp.iucr
fp.ward
< Back
Next >
Cancel

Configure Import Settings

Detected file format: csv

Encoding: utf-8

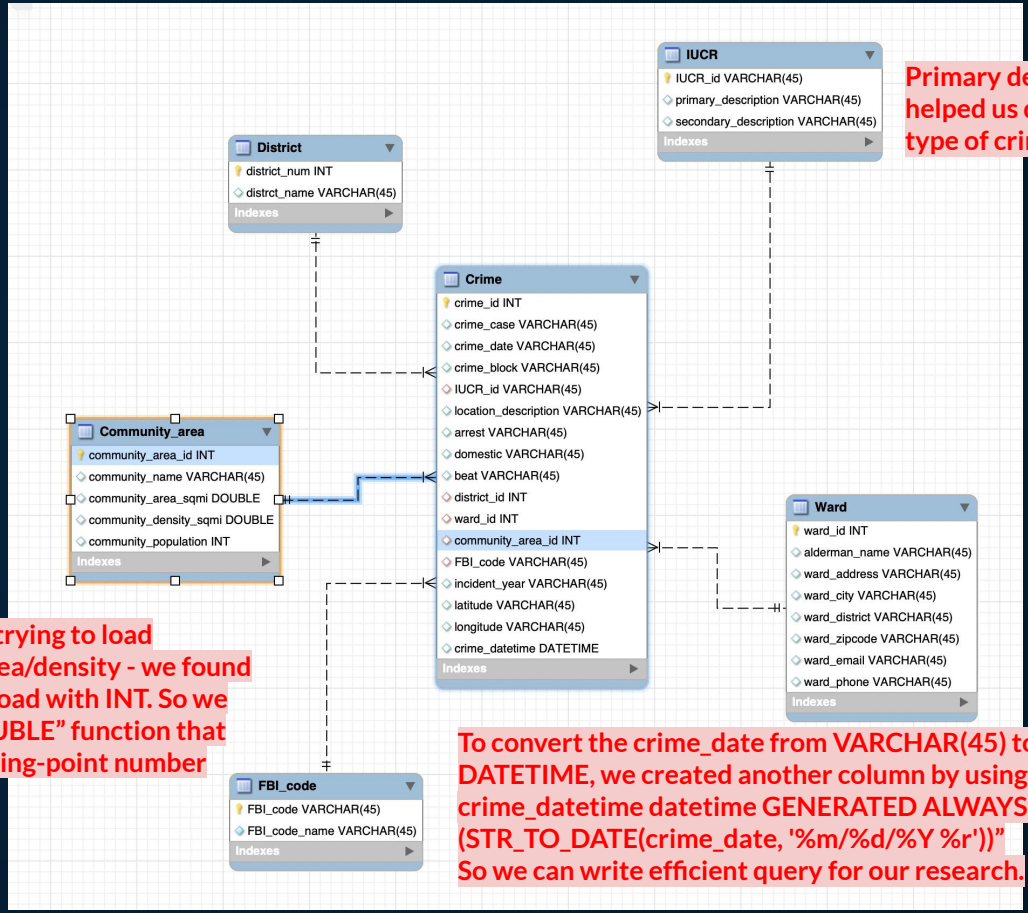
Source Column	Dest Column
ID	crime_id
Case Number	crime_case
Date	crime_date
Block	crime_block
IUCR	IUCR_id

ID	Case Number	Date	Block	IUCR	Location De...	Arrest	Domestic	B
12345411	JE205618	01/01/202...	036XX S...	1320	PARKING...	FALSE	FALSE	
12449065	JE319016	01/01/202...	100XX S...	1750	RESIDEN...	FALSE	TRUE	

< Back Next > Cancel

EER

- There are six tables
  - Crime
  - Community_area
  - Ward
  - FBI_code
  - IUCR
  - District



Primary description helped us define what type of crime it is

When we are trying to load community area/density - we found that we can't load with INT. So we used the "DOUBLE" function that returns a floating-point number

To convert the crime_date from VARCHAR(45) to DATETIME, we created another column by using " crime_datetime datetime GENERATED ALWAYS AS (STR_TO_DATE(crime_date, '%m/%d/%Y %r'))" So we can write efficient query for our research.

SQL QUERY

OVERALL DATA

```sql
-- overall data 200139
SELECT
    COUNT(a.crime_case)
FROM
    crime a
        JOIN
    Community_area b ON a.community_area_id = b.community_area_id
        JOIN
    district c ON a.district_id = c.district_num
        JOIN
    FBI_code d ON a.FBI_code = d.FBI_code
        JOIN
    IUCR e ON a.IUCR_id = e.IUCR_id
        JOIN
    Ward f ON a.ward_id = f.ward_id;
```

We have 200K rows of data

```sql
crime_datetime >= '2021-01-01 00:00:00' AND a.crime_datetime <= '2021-12-31 23:59:59'
```

SQL QUERY

Overall crimes in 2021 per month

October has most numbers of crimes in 2021

```sql
SELECT
    EXTRACT(MONTH FROM a.crime_datetime) AS month,
    COUNT(a.crime_case)
FROM
    crime a
        JOIN
    Community_area b ON a.community_area_id = b.community_area_id
        JOIN
    district c ON a.district_id = c.district_num
        JOIN
    FBI_code d ON a.FBI_code = d.FBI_code
        JOIN
    IUCR e ON a.IUCR_id = e.IUCR_id
        JOIN
    Ward f ON a.ward_id = f.ward_id
GROUP BY EXTRACT(MONTH FROM a.crime_datetime)
ORDER BY COUNT(a.crime_case) DESC;
```

month	COUNT(a.crime_case)
10	18594
7	18493
9	18463
6	18054
8	17807
5	17035
11	16783
12	16608
1	15453
3	15382
4	14936
2	12531

SQL QUERY

Overall crimes in each community in 2021

```sql
SELECT
    COUNT(a.crime_case),
    b.community_name,
    (COUNT(a.crime_case) / 200139) AS numbers_of_crime_percentage
FROM
    crime a
        JOIN
    Community_area b ON a.community_area_id = b.community_area_id
        JOIN
    district c ON a.district_id = c.district_num
        JOIN
    FBI_code d ON a.FBI_code = d.FBI_code
        JOIN
    IUCR e ON a.IUCR_id = e.IUCR_id
        JOIN
    Ward f ON a.ward_id = f.ward_id
WHERE
    a.crime_datetime >= '2021-01-01 00:00:00'
        AND a.crime_datetime <= '2021-12-31 23:59:59'
GROUP BY b.community_name
ORDER BY COUNT(a.crime_case) DESC;
```

The Austin community has the most crimes in 2021, 5% of all crimes in the City of Chicago happened here.

count(a.crime_cas...	community_name	numbers_of_crime_percenta...
11230	Austin	0.0561
8223	Near North Side	0.0411
7123	South Shore	0.0356
6789	Near West Side	0.0339
6293	North Lawndale	0.0314
5893	Auburn Gresham	0.0294
5824	Humboldt Park	0.0291
5559	West Town	0.0278
5495	(The) Loop[11]	0.0275
5448	Greater Grand Crossing	0.0272
5225	Roseland	0.0261
5221	Chatham	0.0261
4744	West Englewood	0.0237
4689	Lake View	0.0234
4561	Englewood	0.0228
4466	Chicago Lawn	0.0223
4121	West Garfield Park	0.0206
3589	East Garfield Park	0.0179
3571	Logan Square	0.0178
3521	South Chicago	0.0176
3438	New City	0.0172
3435	West Pullman	0.0172
3307	Rogers Park	0.0165
3228	South Lawndale	0.0161
3143	Uptown	0.0157
2975	Lincoln Park	0.0149
2958	West Ridge	0.0148
2752	Woodlawn	0.0138
2705	Grand Boulevard	0.0135
2486	Washington Heights	0.0124
2432	Edgewater	0.0122
2248	Portage Park	0.0112
2225	Douglas	0.0111
2199	Irving Park	0.0110
2175	Lower West Side	0.0109
2152	Belmont Cragin	0.0108
1886	Washington Park	0.0094
1834	Albany Park	0.0092
1787	Near South Side	0.0089
1775	Ashburn	0.0089
1764	Brighton Park	0.0088
1713	Morgan Park	0.0086
1680	Hyde Park	0.0084
1673	Lincoln Square	0.0084
1650	Gage Park	0.0082

SQL QUERY

Classify the community & add percentile

```sql
WITH s AS (
    SELECT
        b.community_name,
        COUNT(*) as num_crimes
    FROM
        crime a
            JOIN
        Community_area b ON a.community_area_id = b.community_area_id
    GROUP BY b.community_area_id
),
p AS (
    SELECT
        community_name,
        PERCENT_RANK() OVER (ORDER BY num_crimes) as percentile -- calculate percentile of community by number of crimes
    FROM s
)
SELECT s.community_name, s.num_crimes, p.percentile,
(CASE WHEN p.percentile < 0.25 THEN "LOW CRIME" WHEN p.percentile < 0.5 THEN "MID-LOW CRIME" WHEN p.percentile < 0.75 THEN "MID-HIGH CRIME"
    ELSE "HIGH CRIME" END) AS crime_classification
FROM
    s JOIN p ON s.community_name = p.community_name
ORDER BY p.percentile DESC;
```

SQL QUERY

There are four classifications of crime and Austin is in High Crime

community_name	num_crimes	percentile	crime_classificati...
Austin	11230	1	HIGH CRIME
Near North Side	8223	0.9868421052631579	HIGH CRIME
South Shore	7123	0.9736842105263158	HIGH CRIME
Near West Side	6789	0.9605263157894737	HIGH CRIME
North Lawndale	6293	0.9473684210526315	HIGH CRIME
Auburn Gresham	5893	0.9342105263157895	HIGH CRIME
Humboldt Park	5824	0.9210526315789473	HIGH CRIME
West Town	5559	0.9078947368421053	HIGH CRIME
(The) Loop[11]	5495	0.8947368421052632	HIGH CRIME
Greater Grand Crossing	5448	0.881578947368421	HIGH CRIME
Roseland	5225	0.868421052631579	HIGH CRIME
Chatham	5221	0.8552631578947368	HIGH CRIME
West Englewood	4744	0.8421052631578947	HIGH CRIME
Lake View	4689	0.8289473684210527	HIGH CRIME
Englewood	4561	0.8157894736842105	HIGH CRIME
Chicago Lawn	4466	0.8026315789473685	HIGH CRIME
West Garfield Park	4121	0.7894736842105263	HIGH CRIME
East Garfield Park	3589	0.7763157894736842	HIGH CRIME
Logan Square	3571	0.7631578947368421	HIGH CRIME
South Chicago	3521	0.75	HIGH CRIME
New City	3438	0.7368421052631579	MID-HIGH CRIME
West Pullman	3435	0.7236842105263158	MID-HIGH CRIME
Rogers Park	3307	0.7105263157894737	MID-HIGH CRIME
South Lawndale	3228	0.6973684210526315	MID-HIGH CRIME
Uptown	3143	0.6842105263157895	MID-HIGH CRIME
Lincoln Park	2975	0.6710526315789473	MID-HIGH CRIME
West Ridge	2958	0.6578947368421053	MID-HIGH CRIME
Woodlawn	2752	0.6447368421052632	MID-HIGH CRIME
Grand Boulevard	2705	0.631578947368421	MID-HIGH CRIME
Washington Heights	2486	0.618421052631579	MID-HIGH CRIME
Edgewater	2432	0.6052631578947368	MID-HIGH CRIME
Portage Park	2248	0.5921052631578947	MID-HIGH CRIME
Douglas	2225	0.5789473684210527	MID-HIGH CRIME
Irving Park	2199	0.5657894736842105	MID-HIGH CRIME
Lower West Side	2175	0.5526315789473685	MID-HIGH CRIME
Belmont Cragin	2152	0.5394736842105263	MID-HIGH CRIME

community_name	num_crimes	percentile	crime_classificati...
Gage Park	1650	0.4210526315789473...	MID-LOW CRIME
Avondale	1564	0.4078947368421053...	MID-LOW CRIME
Garfield Ridge	1493	0.3947368421052631...	MID-LOW CRIME
South Deering	1407	0.381578947368421	MID-LOW CRIME
Kenwood	1295	0.3684210526315789	MID-LOW CRIME
Calumet Heights	1278	0.3552631578947368...	MID-LOW CRIME
Dunning	1259	0.3421052631578947...	MID-LOW CRIME
West Lawn	1247	0.3289473684210526...	MID-LOW CRIME
Bridgeport	1246	0.3157894736842105	MID-LOW CRIME
Norwood Park	1169	0.3026315789473684	MID-LOW CRIME
O'Hare	1160	0.2894736842105263	MID-LOW CRIME
Hermosa	1138	0.2763157894736842...	MID-LOW CRIME
Riverdale	1095	0.2631578947368421	MID-LOW CRIME
Avalon Park	1085	0.25	MID-LOW CRIME
North Center	1014	0.2368421052631578...	LOW CRIME
East Side	983	0.2236842105263158	LOW CRIME
Clearing	914	0.2105263157894736...	LOW CRIME
Pullman	906	0.1842105263157894...	LOW CRIME
Armour Square	906	0.1842105263157894...	LOW CRIME
Jefferson Park	828	0.1710526315789473...	LOW CRIME
Beverly	787	0.1578947368421053...	LOW CRIME
West Elsdon	696	0.1447368421052631...	LOW CRIME
McKinley Park	693	0.1315789473684211...	LOW CRIME
North Park	689	0.11842105263157894	LOW CRIME
Archer Heights	663	0.1052631578947368...	LOW CRIME
Hegewisch	596	0.0921052631578947...	LOW CRIME
Oakland	569	0.0789473684210526...	LOW CRIME
Fuller Park	542	0.0657894736842105...	LOW CRIME
Mount Greenwood	499	0.0526315789473684...	LOW CRIME
Forest Glen	465	0.0394736842105263...	LOW CRIME
Burnside	306	0.0263157894736842...	LOW CRIME
Edison Park	240	0.0131578947368421...	LOW CRIME
Montclare	202	0	LOW CRIME

SQL QUERY

Overall crimes in each district in 2021

```
SELECT
    COUNT(a.crime_case),
    c.distrct_name,
    (COUNT(a.crime_case) / 200139) AS numbers_of_crime_percentage
FROM
    crime a
        JOIN
    Community_area b ON a.community_area_id = b.community_area_id
        JOIN
    district c ON a.district_id = c.district_num
        JOIN
    FBI_code d ON a.FBI_code = d.FBI_code
        JOIN
    IUCR e ON a.IUCR_id = e.IUCR_id
        JOIN
    Ward f ON a.ward_id = f.ward_id
WHERE
    a.crime_datetime >= '2021-01-01 00:00:00'
        AND a.crime_datetime <= '2021-12-31 23:59:59'
GROUP BY c.distrct_name
ORDER BY COUNT(a.crime_case) DESC;
```

The city's top 3 most dangerous districts lie on the west side and south side respectively.

COUNT(a.crime_case)	distrct_name	numbers_of_crime_percenta...
13259	11TH	0.0662
13222	6TH	0.0661
12501	8TH	0.0625
11890	4TH	0.0594
10687	12TH	0.0534
10164	3RD	0.0508
10095	7TH	0.0504
9969	18TH	0.0498
9530	19TH	0.0476
9336	2ND	0.0466
9307	5TH	0.0465
8985	1ST	0.0449
8888	9TH	0.0444
8857	10TH	0.0443
8709	25TH	0.0435
7658	15TH	0.0383
6925	16TH	0.0346
6834	22ND	0.0341
6757	24TH	0.0338
6753	14TH	0.0337
5644	17TH	0.0282
4160	20TH	0.0208
9	31ST	0.0000

SQL QUERY

Classify the district & add percentile

```sql
WITH s AS (
    SELECT
        c.distrct_name as district_name,
        COUNT(*) as num_crimes
    FROM
        crime a
            JOIN
        Community_area b ON a.community_area_id = b.community_area_id
            JOIN
        district c ON a.district_id = c.district_num
    GROUP BY district_name
),
p AS (
    SELECT
        district_name,
        PERCENT_RANK() OVER (ORDER BY num_crimes) as percentile -- calculate percentile of community by number of crimes
    FROM s
)
SELECT s.district_name, s.num_crimes, p.percentile,
(CASE WHEN p.percentile < 0.25 THEN "LOW CRIME" WHEN p.percentile < 0.5 THEN "MID-LOW CRIME" WHEN p.percentile < 0.75 THEN "MID-HIGH CRIME"
    ELSE "HIGH CRIME" END) AS crime_classification
FROM
    s JOIN p ON s.district_name = p.district_name
ORDER BY p.percentile DESC;
```

SQL QUERY

Districts with the highest crimes

district_name	num_crimes	percentile	crime_classificati...
11TH	13259	1	HIGH CRIME
6TH	13222	0.9545454545454546	HIGH CRIME
8TH	12501	0.9090909090909091	HIGH CRIME
4TH	11890	0.8636363636363636	HIGH CRIME
12TH	10687	0.8181818181818182	HIGH CRIME
3RD	10164	0.7727272727272727	HIGH CRIME
7TH	10095	0.7272727272727273	MID-HIGH CRIME
18TH	9969	0.6818181818181818	MID-HIGH CRIME
19TH	9530	0.6363636363636364	MID-HIGH CRIME
2ND	9336	0.5909090909090909	MID-HIGH CRIME
5TH	9307	0.5454545454545454	MID-HIGH CRIME
1ST	8985	0.5	MID-HIGH CRIME
9TH	8888	0.454545454545454...	MID-LOW CRIME
10TH	8857	0.4090909090909091	MID-LOW CRIME
25TH	8709	0.363636363636363...	MID-LOW CRIME
15TH	7658	0.3181818181818182	MID-LOW CRIME
16TH	6925	0.2727272727272727	MID-LOW CRIME
22ND	6834	0.227272727272727...	LOW CRIME
24TH	6757	0.181818181818181...	LOW CRIME
14TH	6753	0.136363636363636...	LOW CRIME
17TH	5644	0.090909090909090...	LOW CRIME
20TH	4160	0.045454545454545...	LOW CRIME
31ST	9	0	LOW CRIME

SQL QUERY

Type of crimes in 2021

```
SELECT
    COUNT(a.crime_case),
    e.IUCR_id,
    e.primary_description,
    (COUNT(a.crime_case) / 200139) AS crime_type_percentage
FROM
    crime a
        JOIN
    Community_area b ON a.community_area_id = b.community_area_id
        JOIN
    district c ON a.district_id = c.district_num
        JOIN
    FBI_code d ON a.FBI_code = d.FBI_code
        JOIN
    IUCR e ON a.IUCR_id = e.IUCR_id
        JOIN
    Ward f ON a.ward_id = f.ward_id
WHERE
    a.crime_datetime >= '2021-01-01 00:00:00'
        AND a.crime_datetime <= '2021-12-31 23:59:59'
GROUP BY e.primary_description
ORDER BY COUNT(a.crime_case) DESC;
```

Through this SQL query we find that theft, battery, and criminal damage are the most common crimes.

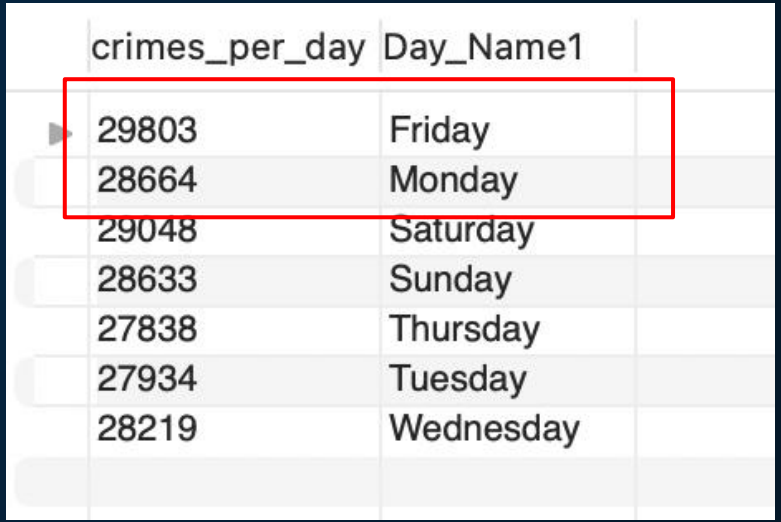
COUNT(a.crime_case)	IUCR_id	primary_description	crime_type_percenta...
40150	810	THEFT	0.2006
36278	460	BATTERY	0.1813
24802	1310	CRIMINAL DAMAGE	0.1239
19684	560	ASSAULT	0.0984
16678	1153	DECEPTIVE PRACTICE	0.0833
13664	2820	OTHER OFFENSE	0.0683
10474	910	MOTOR VEHICLE THEFT	0.0523
8863	143A	WEAPONS VIOLATION	0.0443
7845	031A	ROBBERY	0.0392
6582	610	BURGLARY	0.0329
4629	2026	NARCOTICS	0.0231
3292	1365	CRIMINAL TRESPASS	0.0164
1528	1780	OFFENSE INVOLVING...	0.0076
1468	281	CRIMINAL SEXUAL ASS...	0.0073
884	1570	SEX OFFENSE	0.0044
802	110	HOMICIDE	0.0040
597	470	PUBLIC PEACE VIOLAT...	0.0030
516	1020	ARSON	0.0026
369	584	STALKING	0.0018
311	3731	INTERFERENCE WITH...	0.0016
182	2250	LIQUOR LAW VIOLATION	0.0009
164	1479	CONCEALED CARRY LI...	0.0008
114	3970	INTIMIDATION	0.0006
95	1512	PROSTITUTION	0.0005
86	1792	KIDNAPPING	0.0004
49	1540	OBSCENITY	0.0002
13	1661	GAMBLING	0.0001
10	1055	HUMAN TRAFFICKING	0.0000
4	1481	NON-CRIMINAL	0.0000
4	1536	PUBLIC INDECENCY	0.0000
2	1900	OTHER NARCOTIC VIO...	0.0000

SQL QUERY

Which weekdays have most crime?

This query shows that criminals commit the most crimes on Friday and Saturday. This makes sense as their may be higher opportunity on the weekends.

```sql
SELECT
    COUNT(a.crime_id) AS crimes_per_day,
    DAYNAME(a.crime_datetime) AS Day_Name1
FROM
    crime a
        JOIN
    Community_area b ON a.community_area_id = b.community_area_id
        JOIN
    district c ON a.district_id = c.district_num
        JOIN
    FBI_code d ON a.FBI_code = d.FBI_code
        JOIN
    IUCR e ON a.IUCR_id = e.IUCR_id
GROUP BY Day_Name1
ORDER BY Day_Name1;
```



crimes_per_day	Day_Name1
29803	Friday
28664	Monday
29048	Saturday
28633	Sunday
27838	Thursday
27934	Tuesday
28219	Wednesday

SQL QUERY

What time are crimes most often

Criminals commit the most crimes around midnight and noon. This may indicate criminals commit more crime around lunch as people are moving around more during their lunch break.

```sql
SELECT
    COUNT(a.crime_id) AS crimes_per_time,
    TIME(a.crime_datetime) AS time
FROM
    crime a
        JOIN
    Community_area b ON a.community_area_id = b.community_area_id
        JOIN
    district c ON a.district_id = c.district_num
        JOIN
    FBI_code d ON a.FBI_code = d.FBI_code
        JOIN
    IUCR e ON a.IUCR_id = e.IUCR_id
GROUP BY TIME(a.crime_datetime)
ORDER BY crimes_per_time DESC;
```

crimes_per_time	time
7140	00:00:00
5271	12:00:00
4205	09:00:00
3478	15:00:00
3463	17:00:00
3396	18:00:00
3287	10:00:00
3170	16:00:00
3133	14:00:00
3118	20:00:00
3075	22:00:00
3035	19:00:00
2985	13:00:00
2959	21:00:00
2681	08:00:00
2595	11:00:00
2346	23:00:00
1890	01:00:00
1818	02:00:00
1594	07:00:00
1575	00:01:00
1515	03:00:00
1501	15:30:00
1436	16:30:00
1408	18:30:00
1405	17:30:00
1354	20:30:00
1352	14:30:00
1340	19:30:00
1335	21:30:00
1333	22:30:00
1298	10:30:00
1294	13:30:00
1248	12:30:00
1232	06:00:00
1167	04:00:00
1163	11:30:00

SQL QUERY

Which day of year has the most crimes?

```
SELECT
    COUNT(a.crime_id) AS crimes_per_day,
    DATE(a.crime_datetime) AS day
FROM
    crime a
        JOIN
    Community_area b ON a.community_area_id = b.community_area_id
        JOIN
    district c ON a.district_id = c.district_num
        JOIN
    FBI_code d ON a.FBI_code = d.FBI_code
        JOIN
    IUCR e ON a.IUCR_id = e.IUCR_id
GROUP BY DATE(a.crime_datetime)
ORDER BY crimes_per_day DESC;
```

crimes_per_day	day
787	2021-10-01
758	2021-07-31
720	2021-08-01
700	2021-01-01
700	2021-09-01
694	2021-07-29
688	2021-06-19
686	2021-09-19
679	2021-06-06
678	2021-10-02
673	2021-07-01
672	2021-12-01
668	2021-05-01
666	2021-06-20
661	2021-11-01
658	2021-10-03
656	2021-07-04
656	2021-09-13
653	2021-06-27
652	2021-08-21
650	2021-06-01
650	2021-08-27
649	2021-06-08
647	2021-12-03
647	2021-09-04
646	2021-09-12
645	2021-09-16
643	2021-05-22
643	2021-05-24
643	2021-09-05
643	2021-06-13
642	2021-09-10
639	2021-06-11
639	2021-07-03

October 1st is the day most crimes were committed in Chicago.

SQL QUERY

What happened that day?

7. Stocks post biggest monthly loss since March 2020

U.S. stocks fell on Thursday, ending the last day of September with the biggest monthly loss since the plunge at the start of the coronavirus pandemic. The S&P finished September down by 4.8 percent, its first monthly loss since January and its biggest decline since March 2020. The S&P 500 fell by 1.2 percent on Thursday. The Dow Jones Industrial Average and the Nasdaq lost 1.6 percent and 0.4 percent, respectively. After months of gains in 2021, Wall Street began tumbling in recent weeks as the highly contagious Delta variant drove a coronavirus surge that disrupted the economic recovery. Futures tied to the three main U.S. averages plunged further early Friday and were down another 0.5 percent several hours before the opening bell.

SQL QUERY

Which types of locations have the most crimes?

We found that streets and apartments have the most crime. Streets hold a majority crime locations with over 25%.

```
SELECT
    COUNT(a.crime_case),
    a.location_description,
    (COUNT(a.crime_case) / 200139) AS numbers_of_crime_percentage
FROM
    crime a
        JOIN
    Community_area b ON a.community_area_id = b.community_area_id
        JOIN
    district c ON a.district_id = c.district_num
        JOIN
    FBI_code d ON a.FBI_code = d.FBI_code
        JOIN
    IUCR e ON a.IUCR_id = e.IUCR_id
        JOIN
    Ward f ON a.ward_id = f.ward_id
WHERE
    a.crime_datetime >= '2021-01-01 00:00:00'
        AND a.crime_datetime <= '2021-12-31 23:59:59'
GROUP BY a.location_description
ORDER BY COUNT(a.crime_case) DESC;
```

COUNT(a.crime_case)	location_description	numbers_of_crime_percenta...
50666	STREET	0.2532
42507	APARTMENT	0.2124
30527	RESIDENCE	0.1525
11580	SIDEWALK	0.0579
6286	PARKING LOT / GARAGE (NON RESIDENTIAL)	0.0314
5336	SMALL RETAIL STORE	0.0267
4730	ALLEY	0.0236
3659	RESTAURANT	0.0183
2954	RESIDENCE - PORCH / HALLWAY	0.0148
2889	GAS STATION	0.0144
2872	COMMERCIAL / BUSINESS OFFICE	0.0144
2816	OTHER (SPECIFY)	0.0141
2650	VEHICLE NON-COMMERCIAL	0.0132
2388	DEPARTMENT STORE	0.0119
2332	RESIDENCE - YARD (FRONT / BACK)	0.0117
2302	RESIDENCE - GARAGE	0.0115
1942	GROCERY FOOD STORE	0.0097
1461	PARK PROPERTY	0.0073
1195	HOTEL / MOTEL	0.0060
1136	CONVENIENCE STORE	0.0057
1124	BAR OR TAVERN	0.0056
1062	CTA TRAIN	0.0053
1058	DRUG STORE	0.0053
968	HOSPITAL BUILDING / GROUNDS	0.0048

SQL QUERY

What's the arrest rate in 2021 overall?

```sql
SELECT
    COUNT(a.arrest = 'TRUE' OR NULL) AS arrest_num,
    COUNT(a.arrest = 'TRUE' OR NULL) / COUNT(*) AS arrest_rate,
    COUNT(*) - COUNT(a.arrest = 'TRUE' OR NULL) AS non_arrest_num,
    (COUNT(*) - COUNT(a.arrest = 'TRUE' OR NULL)) / COUNT(*) AS non_arrest_rate
FROM
    crime a
        JOIN
    Community_area b ON a.community_area_id = b.community_area_id
        JOIN
    district c ON a.district_id = c.district_num
        JOIN
    FBI_code d ON a.FBI_code = d.FBI_code
        JOIN
    IUCR e ON a.IUCR_id = e.IUCR_id
        JOIN
    Ward f ON a.ward_id = f.ward_id
WHERE
    a.crime_datetime >= '2021-01-01 00:00:00'
        AND a.crime_datetime <= '2021-12-31 23:59:59'
ORDER BY COUNT(a.arrest) DESC;
```

Arrest rate in Chicago is actually pretty low at 11%.

arrest_num	arrest_rate	non_arrest_num	non_arrest_rate
23887	0.1194	176252	0.8806

SQL QUERY

Which community has the highest arrest rates?

```sql
SELECT
    b.community_name,
    COUNT(a.arrest = 'TRUE' OR NULL) / COUNT(*) as arrest_rate_in_community
FROM
    crime a
        JOIN
    Community_area b ON a.community_area_id = b.community_area_id
        JOIN
    district c ON a.district_id = c.district_num
        JOIN
    FBI_code d ON a.FBI_code = d.FBI_code
        JOIN
    IUCR e ON a.IUCR_id = e.IUCR_id
        JOIN
    Ward f ON a.ward_id = f.ward_id
WHERE
    a.crime_datetime >= '2021-01-01 00:00:00'
        AND a.crime_datetime <= '2021-12-31 23:59:59'
GROUP BY b.community_name
ORDER BY COUNT(a.arrest) DESC;
```

Austin, Near North Side, and South Shore have the highest arrest rate. This makes sense as they also have highest count of crimes.

community_name	arrest_rate_in_commun...
Austin	0.1541
Near North Side	0.1216
South Shore	0.0856
Near West Side	0.0875
North Lawndale	0.2140
Auburn Gresham	0.1420
Humboldt Park	0.2366
West Town	0.0718
(The) Loop[11]	0.1387
Greater Grand Crossing	0.1178
Roseland	0.1380
Chatham	0.1084
West Englewood	0.1408
Lake View	0.0855
Englewood	0.1432
Chicago Lawn	0.1220
West Garfield Park	0.2975
East Garfield Park	0.2159
Logan Square	0.0891
South Chicago	0.0861
New City	0.1358
West Pullman	0.1202
Rogers Park	0.1031
South Lawndale	0.1314
Uptown	0.0904
Lincoln Park	0.0491
West Ridge	0.0757
Woodlawn	0.1047
Grand Boulevard	0.0939

SQL QUERY

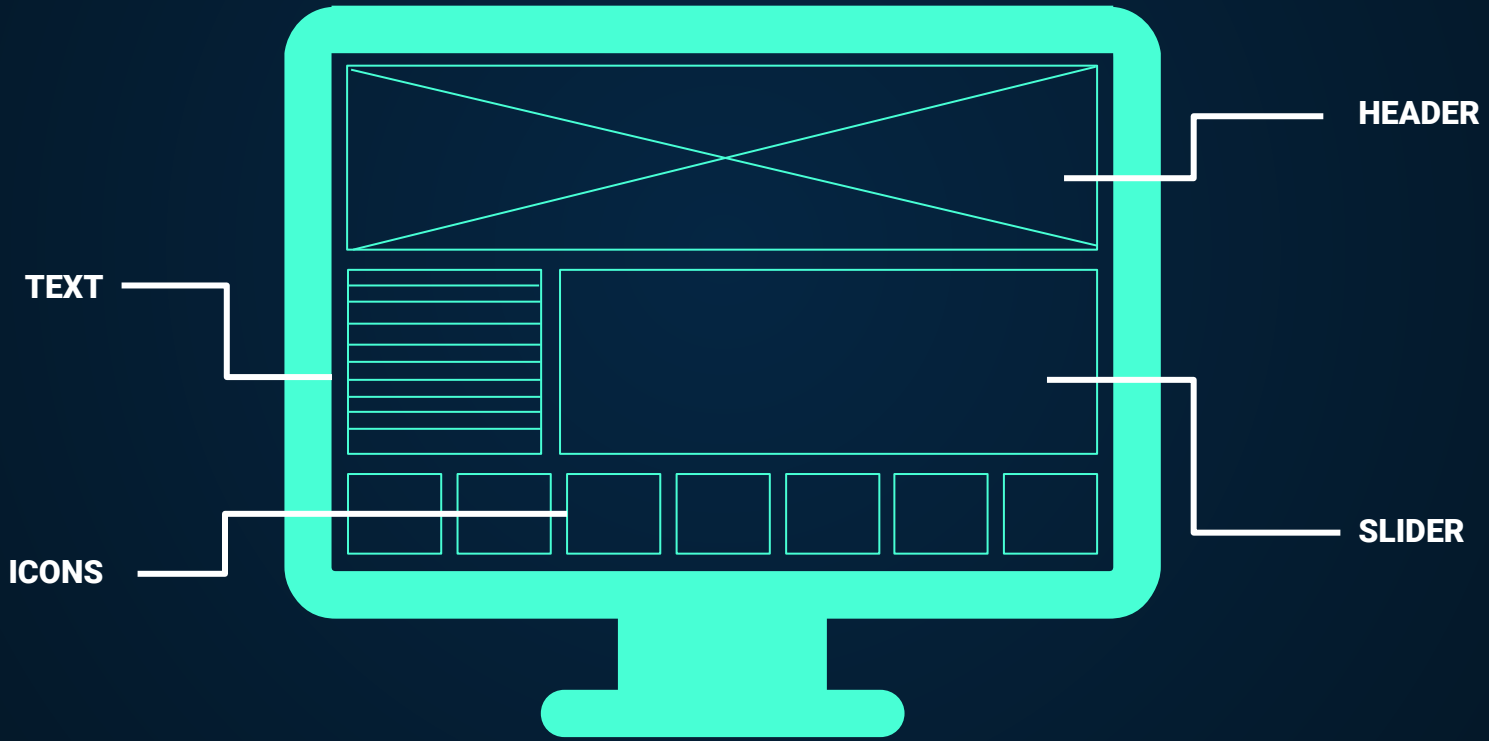
Do higher populations equate to more crimes in Chicago?

There is no correlation between count of population and count of crimes.

```sql
SELECT
    b.community_name,
    b.community_population,
    COUNT(*),
    b.community_population / COUNT(*) AS per_count_of_population
FROM
    crime a
        JOIN
    Community_area b ON a.community_area_id = b.community_area_id
        JOIN
    district c ON a.district_id = c.district_num
        JOIN
    FBI_code d ON a.FBI_code = d.FBI_code
        JOIN
    IUCR e ON a.IUCR_id = e.IUCR_id
        JOIN
    Ward f ON a.ward_id = f.ward_id
WHERE
    a.crime_datetime >= '2021-01-01 00:00:00'
        AND a.crime_datetime <= '2021-12-31 23:59:59'
GROUP BY b.community_name
ORDER BY b.community_population DESC;
```

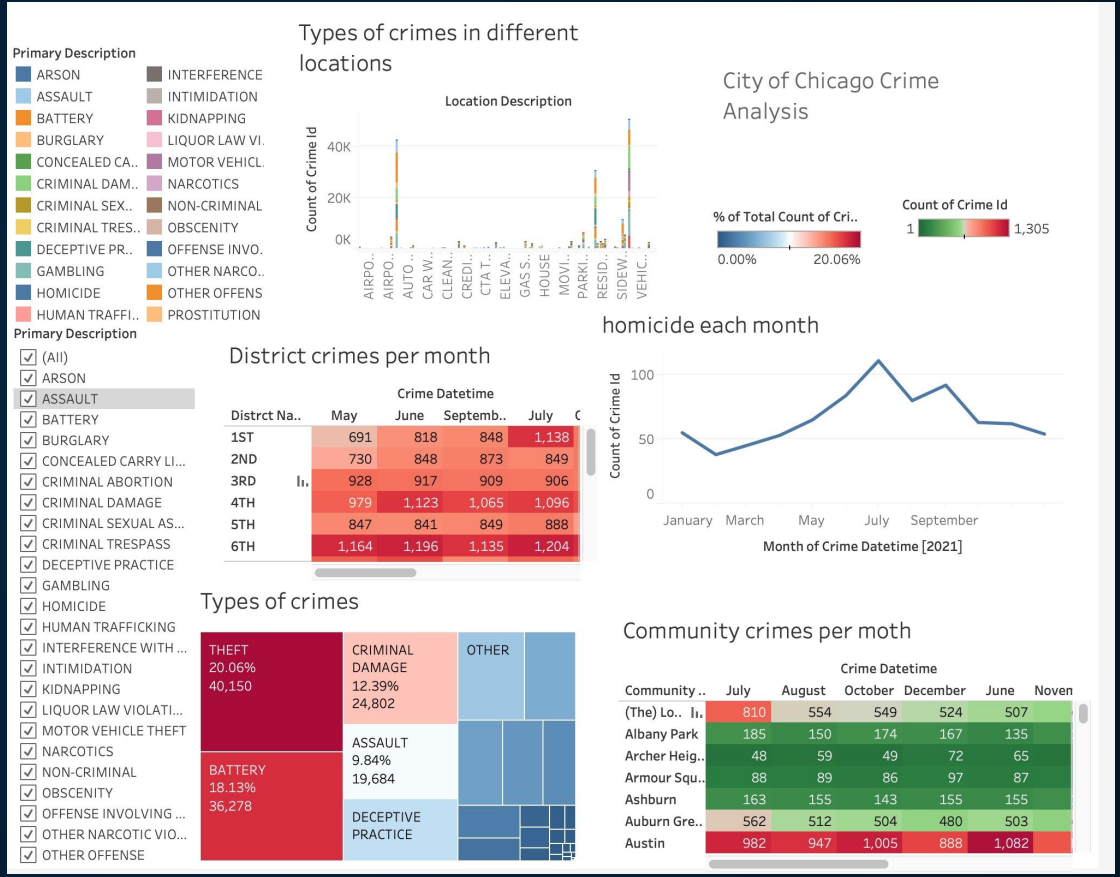
community_name	community_populati...	COUNT(*)	per_count_of_population
Near North Side	105481	8223	12.8276
Lake View	103050	4689	21.9770
Austin	96557	11230	8.5981
West Town	87781	5559	15.7908
Belmont Cragin	78116	2152	36.2993
West Ridge	77122	2958	26.0723
Logan Square	71665	3571	20.0686
South Lawndale	71399	3228	22.1186
Lincoln Park	70492	2975	23.6948
Near West Side	67881	6789	9.9987
Portage Park	63020	2248	28.0338
Uptown	57182	3143	18.1934
Edgewater	56296	2432	23.1480
Chicago Lawn	55931	4466	12.5237
Rogers Park	55628	3307	16.8213
Humboldt Park	54165	5824	9.3003
South Shore	53971	7123	7.5770
Irving Park	51940	2199	23.6198
Albany Park	48396	1834	26.3882
Brighton Park	45053	1764	25.5402
Auburn Gresham	44878	5893	7.6155
New City	43628	3438	12.6899
Dunning	43147	1259	34.2708
(The) Loop[11]	42298	5495	7.6975
Ashburn	41098	1775	23.1538
Lincoln Square	40494	1673	24.2044
Gage Park	39540	1650	23.9636
Roseland	38816	5225	7.4289
Norwood Park	38303	1169	32.7656
Avondale	36257	1564	23.1822
Garfield Ridge	35439	1493	23.7368
North Center	35114	1014	34.6292
North Lawndale	34794	6293	5.5290

VISUALIZATION



VISUALIZATION

TABLEAU



VISUALIZATION

Community crimes per moth

Crime Datetime

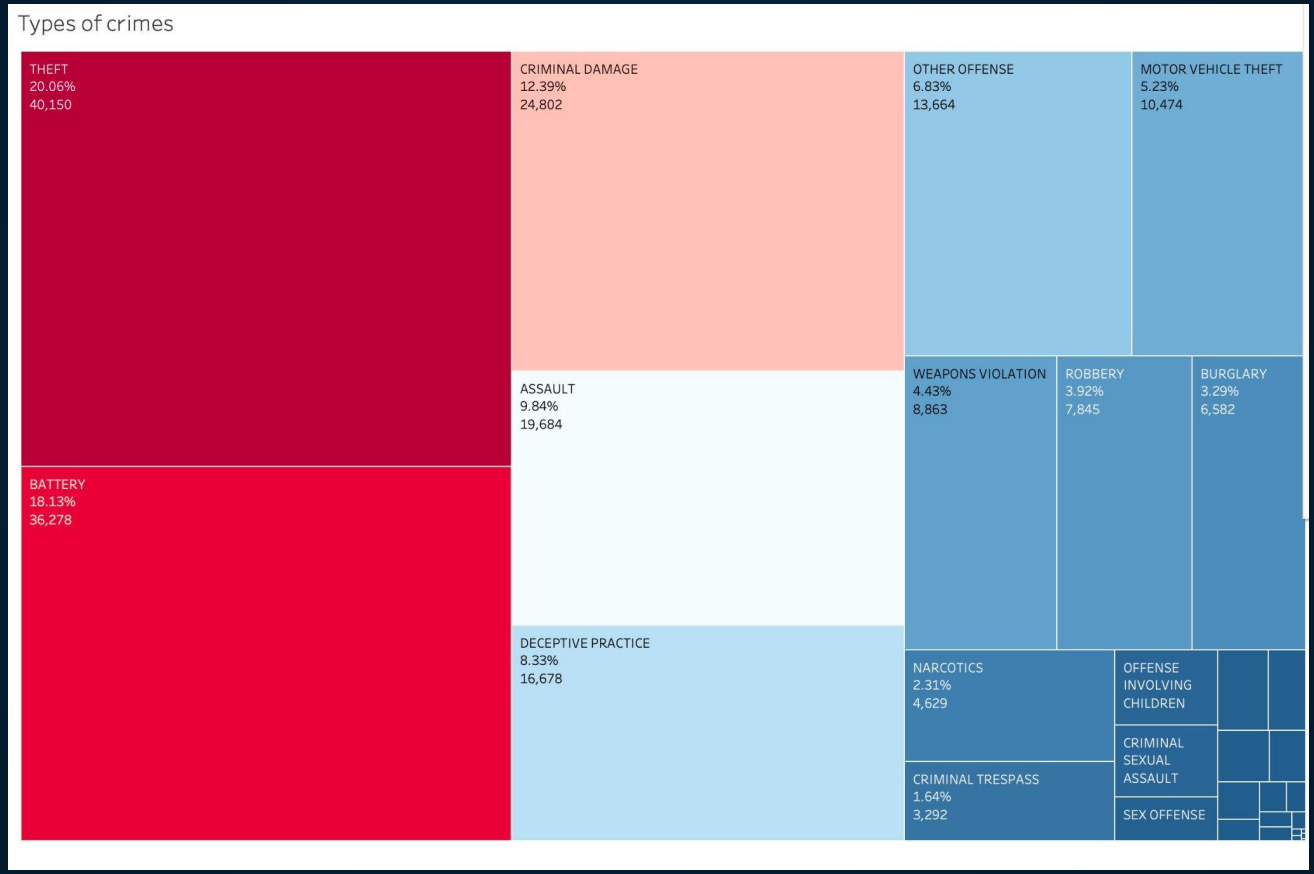
Community Name	January	February	March	April	May	June	July	August	Septemb..	October	November	December
(The) Loop[11]	304	239	257	344	431	507	810	554	487	549	489	524
Albany Park	144	113	128	135	157	135	185	150	169	174	177	167
Archer Heights	53	42	50	43	52	65	48	59	72	49	58	72
Armour Square	65	48	47	48	61	87	88	89	101	86	89	97
Ashburn	145	114	111	147	149	155	163	155	167	143	171	155
Auburn Gresham	492	383	531	459	516	503	562	512	483	504	468	480
Austin	863	718	987	884	1,030	1,082	982	947	1,001	1,005	843	888
Avalon Park	98	57	79	81	81	123	104	90	90	109	91	82
Avondale	125	106	136	109	120	129	127	140	145	152	140	135
Belmont Cragin	197	138	161	170	181	168	196	183	169	198	195	196
Beverly	70	52	55	57	56	67	86	62	60	81	76	65
Bridgeport	124	102	82	80	90	101	127	121	113	112	99	95
Brighton Park	138	112	124	136	151	139	142	159	157	185	165	156
Burnside	29	17	25	28	14	26	23	29	27	35	24	29
Calumet Heights	132	74	100	91	89	121	108	109	115	129	105	105
Chatham	405	337	414	392	446	494	451	468	521	491	407	395
Chicago Lawn	387	321	355	343	366	369	384	398	409	393	377	364
Clearing	77	69	70	62	60	63	84	96	101	82	70	80
Douglas	163	125	205	173	179	215	217	197	186	193	187	185
Dunning	112	86	92	85	114	117	117	121	125	88	102	100
East Garfield Park	270	234	308	309	318	351	351	323	338	316	253	218
East Side	72	50	85	91	86	95	92	102	87	87	63	73
Edgewater	187	177	177	156	189	214	237	192	206	224	243	230
Edison Park	23	16	13	13	20	17	28	27	19	19	26	19
Englewood	369	245	352	401	436	483	431	396	386	400	346	316
Forest Glen	51	37	30	28	41	51	51	33	40	25	41	37
Fuller Park	39	27	42	38	47	44	54	62	48	41	53	47
Gage Park	124	99	126	126	160	147	152	149	142	156	131	138
Garfield Ridge	132	102	131	98	119	140	125	121	143	143	115	124
Grand Boulevard	206	192	198	184	205	259	230	231	254	260	243	243
Greater Grand Crossing	429	379	467	421	516	476	496	447	472	457	474	414
Hegewisch	38	44	52	37	38	49	57	59	68	60	39	55
Hermosa	78	70	77	67	89	97	104	96	133	112	105	110
Humboldt Park	439	390	517	501	558	545	513	491	457	476	499	438
Hyde Park	103	83	97	103	131	151	153	167	183	189	170	150
Irving Park	167	140	153	163	154	156	221	200	197	242	219	187
Jefferson Park	93	62	61	69	52	70	69	73	87	78	60	54
Kenwood	92	77	84	90	114	119	117	122	123	121	116	120

VISUALIZATION

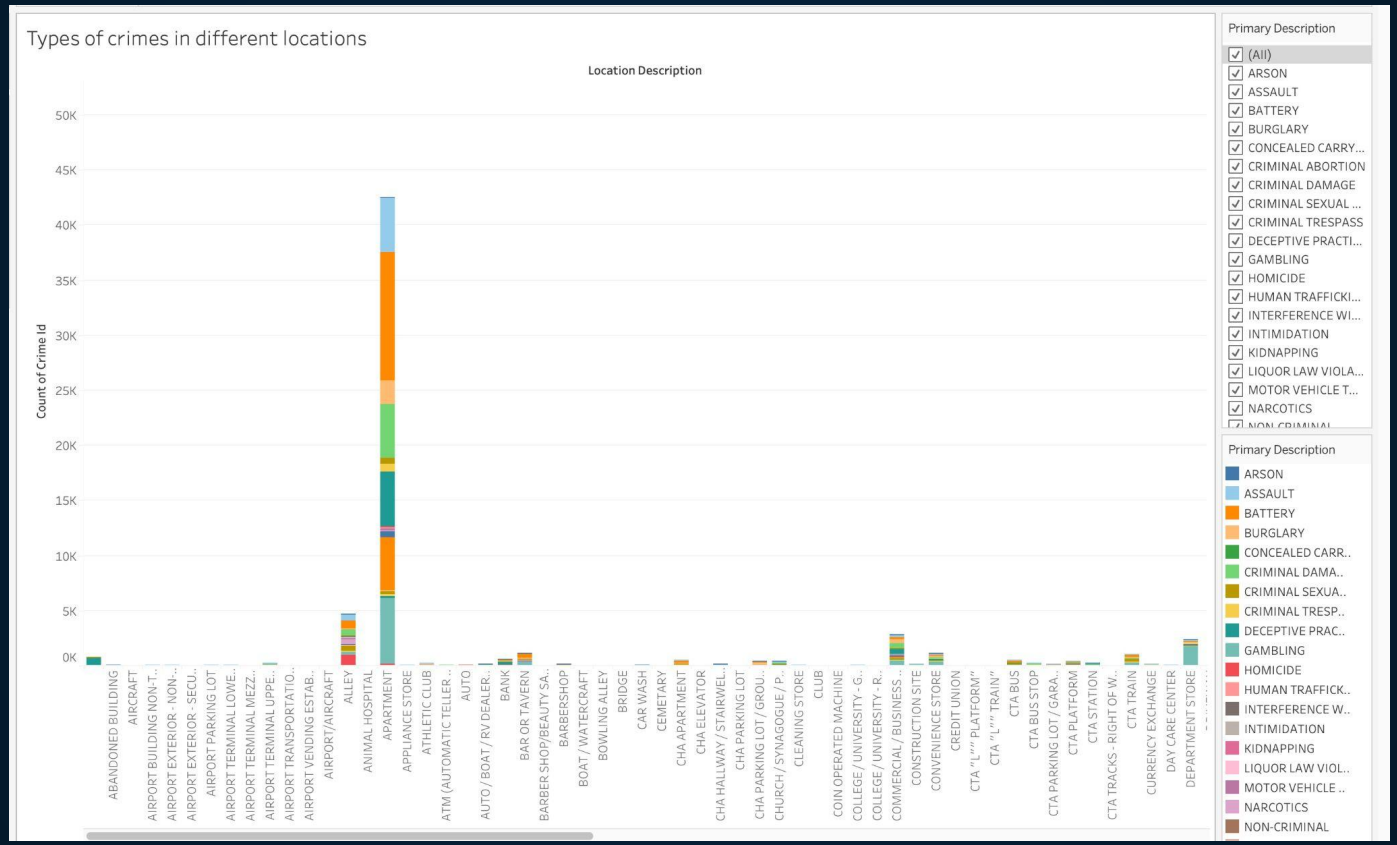
district crimes per month

	Crime Datetime											
Distrct Na..	January	February	March	April	May	June	July	August	Septemb..	October	November	December
1ST	516	447	504	579	691	818	1,138	861	848	903	819	861
2ND	690	566	687	652	730	848	849	846	873	891	877	827
3RD	858	648	863	763	928	917	906	848	909	879	819	826
4TH	930	742	961	962	979	1,123	1,096	1,106	1,065	1,027	905	994
5TH	778	581	749	739	847	841	888	808	849	769	725	733
6TH	1,042	857	1,164	1,015	1,164	1,196	1,204	1,181	1,135	1,149	1,062	1,053
7TH	788	596	783	895	975	1,049	963	895	844	888	721	698
8TH	1,037	857	972	919	1,030	1,057	1,109	1,113	1,176	1,092	1,055	1,084
9TH	684	571	627	649	729	773	827	832	835	860	769	732
10TH	655	580	688	670	820	851	788	762	846	809	668	720
11TH	1,032	997	1,259	1,204	1,278	1,305	1,147	1,105	1,132	1,057	881	862
12TH	745	625	734	671	876	897	947	908	1,076	1,223	1,015	970
14TH	509	392	565	462	565	574	605	585	635	715	619	527
15TH	588	447	686	625	722	706	689	646	685	692	568	604
16TH	698	495	527	487	576	628	620	598	645	552	547	552
17TH	450	368	411	401	439	425	529	505	517	559	549	491
18TH	620	483	562	549	741	882	999	1,060	1,111	1,097	943	922
19TH	713	498	602	648	741	827	928	939	931	966	891	846
20TH	312	271	316	319	337	341	357	368	393	439	373	334
22ND	598	478	500	564	561	609	624	537	596	633	585	549
24TH	548	434	508	501	543	602	523	604	613	599	613	669
25TH	661	598	714	661	763	783	755	699	749	794	779	753
31ST	1			1		2	2	1		1		1

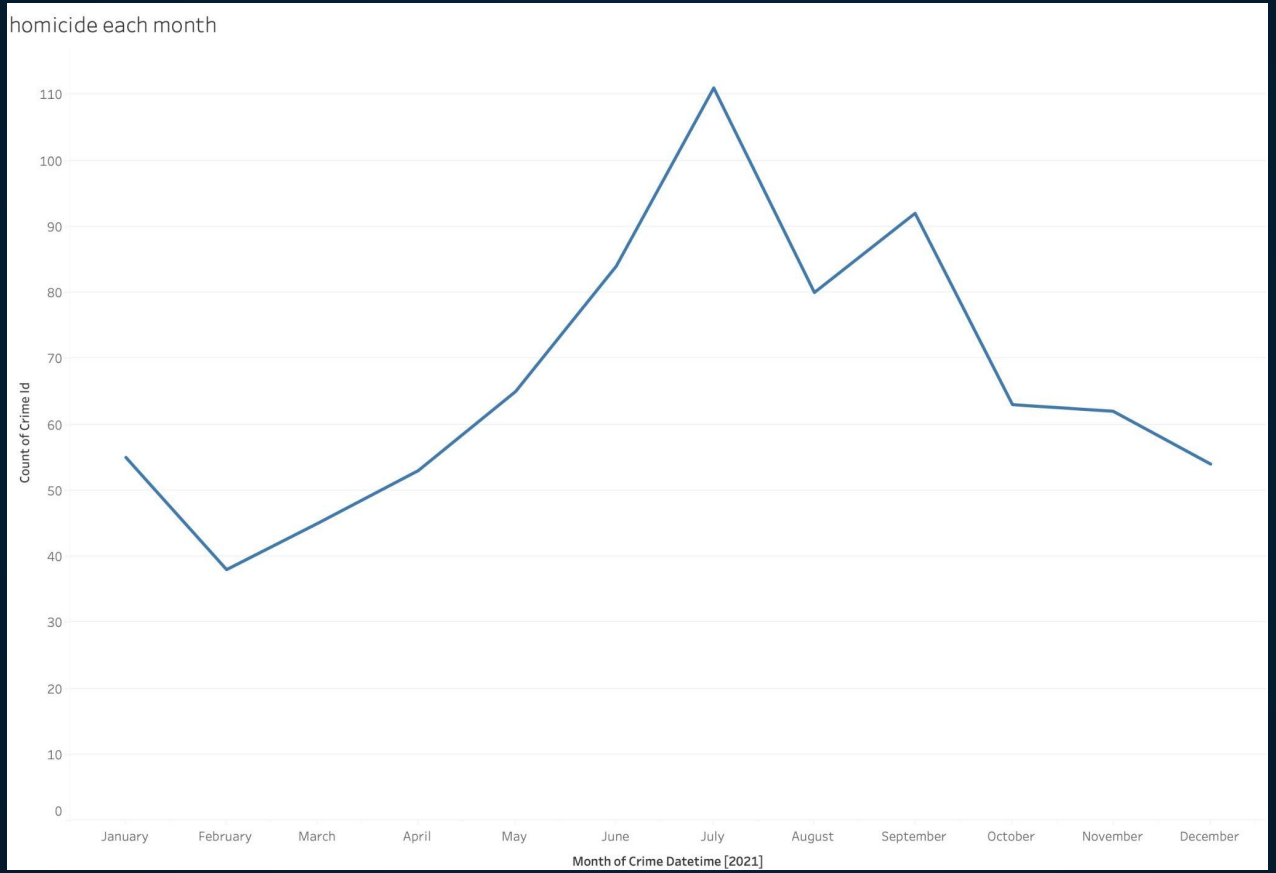
VISUALIZATION



VISUALIZATION



VISUALIZATION

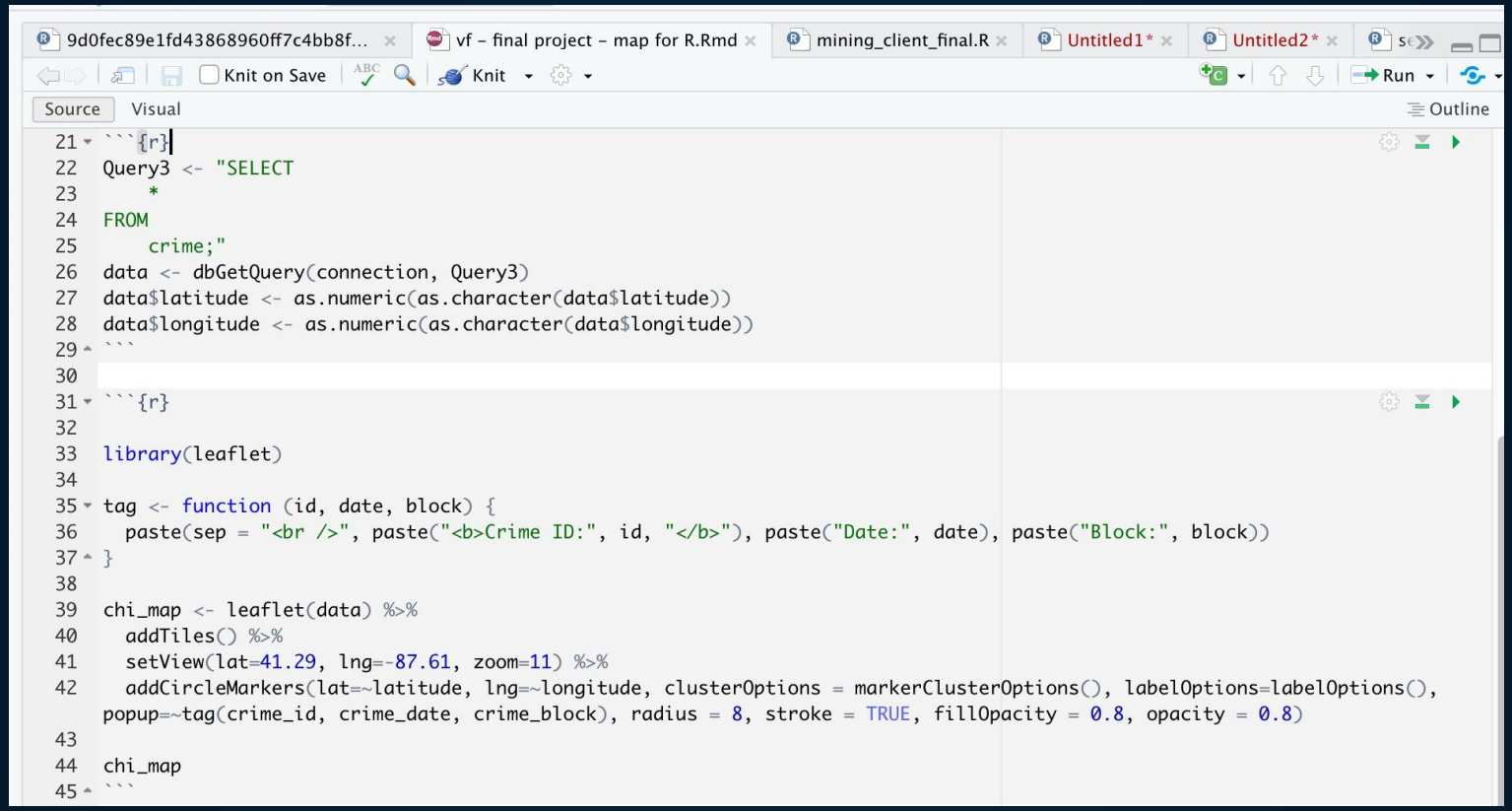


VISUALIZATION

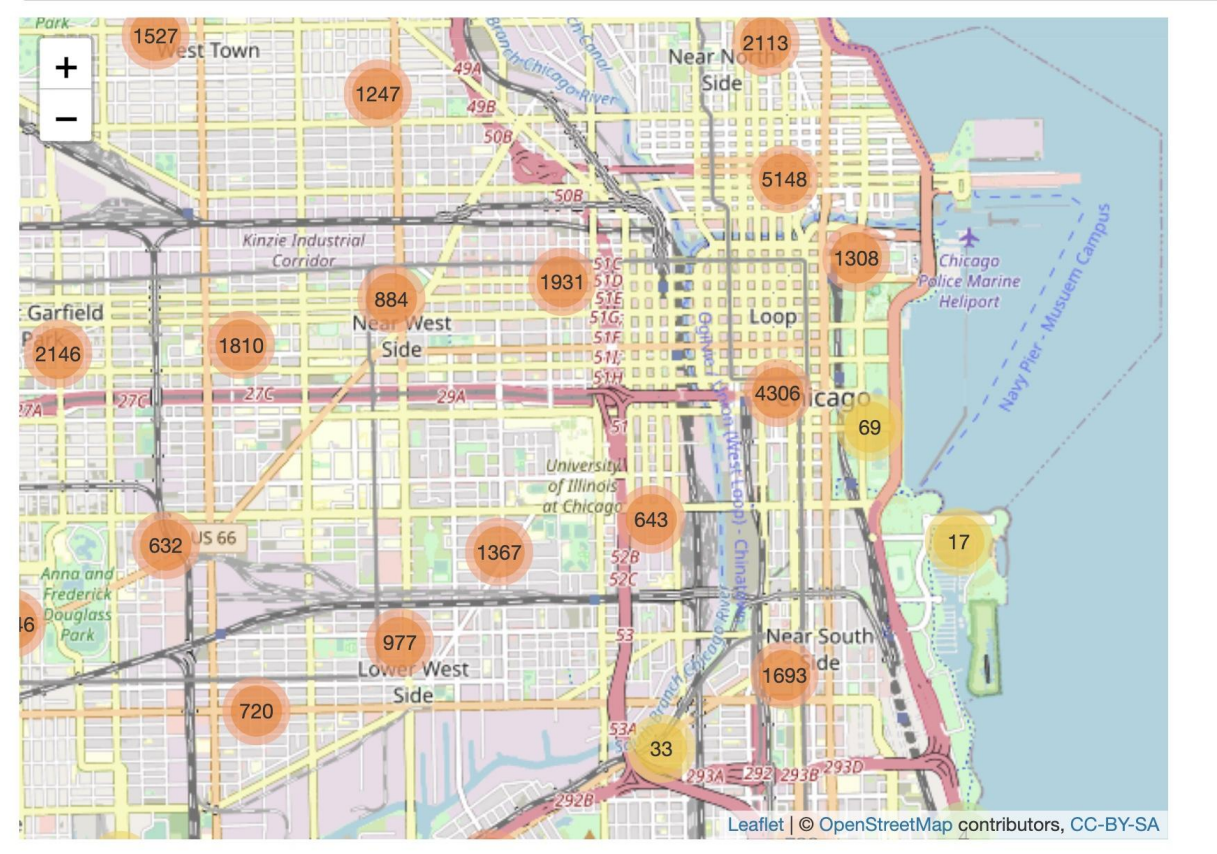
R STUDIO

````
2022/11/29
final project
Zoey Chen

Initialize libraries
```{r, echo=FALSE}
library(RMySQL);
```

Set up the connection
```{r, echo=FALSE}
connection = dbConnect(MySQL(),user="root", password="root",
                       dbname="FP", host="localhost");
```
````

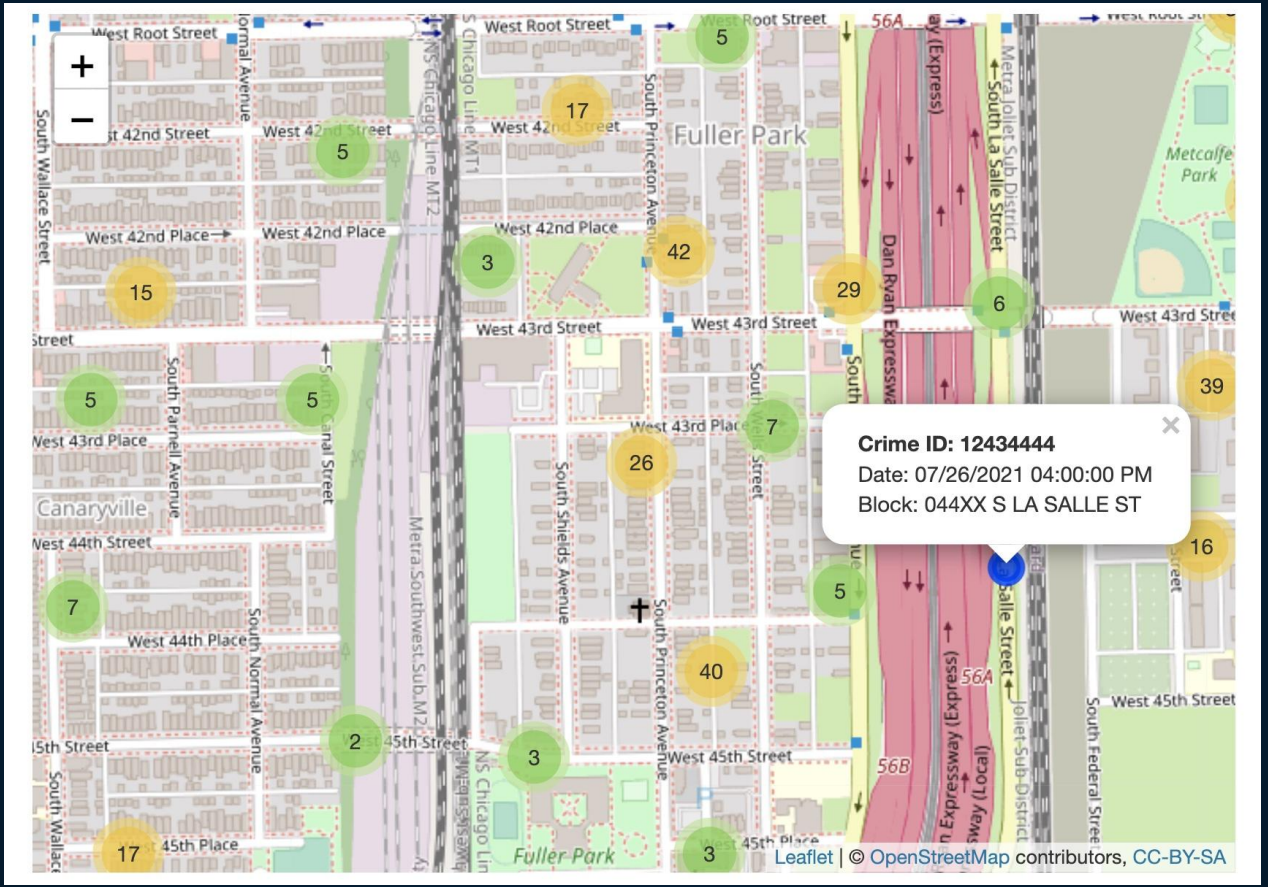
VISUALIZATION



VISUALIZATION



VISUALIZATION



Business Use Case & Recommendations

Provide insights to how crimes in Chicago changes depending on area and time.

Which areas are trending upwards to provide policy recommendations on policing and resource allocation to local government.

In addition our analysis can serve as a guide for Chicago residents on which areas to avoid Recommendations.

Lessons Learned & Improvements

ADAPT CHANGING SITUATIONS

Learned to be more responsible for the tasks and time management.

PROGRAMMING TOOLS PYTHON, MYSQL, R

Learned from our fails to research methods to fix our queries and codes.

KNOWLEDGE ABOUT DATASET

Explore more about the value in the dataset better. Such as the differences between reporting the case on “Street” vs “Sidewalk

THANK YOU

Q&A