A Glimpse into the Patterns in Crime in Los Angeles

Team: Irene Na, Vickram Ramswamy

Introduction:

In recent years, residents of California and elsewhere across the U.S. have expressed concern that the crime rate is increasing, particularly since the outbreak of the COVID-19 pandemic. To investigate the validity of this perception, we will analyze a dataset provided by data.gov containing crime incident data for the City of Los Angeles starting from January, 2020 to March, 2024.

Research Objective:

1. Identify and descriptively test the hypotheses that are related to the public perception regarding the changes of crimes in Los Angeles in recent years.
2. Develop outcome based questions that will help describe patterns, trends & insights that can help LA's law enforcement agencies.

Hypotheses:

1. Hypothesis One - petty crimes in Los Angeles, defined as those with an economic loss of less than $950, increased meaningfully in recent years. - Does our data support this?
2. Hypothesis Two - burglaries from vehicles increased meaningfully in recent years. - Does our data support this?

Outcome Based Questions/descriptive data exploration:

1. Any patterns associated with the time of the crimes?
2. What are the major crime types and the YoY trend of total crime by area?
3. What are the highest crimes and their YoY trend, by area?
4. Which areas have seen the highest number of crimes with firearms?

Details of The Dataset:

- Source: Data.gov
- Dataset: City of Los Angeles - Crime Data from 2020 to Present
    - Data Dictionary
- Data Coverage: Crime incident log from January 2020 to March 4th, 2024 by LAPD
- Total entries: There are 910,707 entries before cleaning.
- Features included:
    - There are 28 features included in total, with descriptions in Data Dictionary
    - More specifically include below categories of features:
        - Case number: Division record number ('DR No.')
        - Time related: date (reported, occurred); time occurred
        - Area related: area code, area name, district number, location, latitude, longitude, premises description (type)
        - Description: crime description (types, with 'petty crime' labeled)
        - Victim Demographics: Age, Gender, Ethnicity
- For our purpose, the most useful dimensional attributes / columns should be:

- Report number, date occurred, time occurred, area name, premises type, crime description, victim age, victim sex, victim descendance.

Exploratory Data Analysis (EDA):

1. Dimension of the data before cleaning: 910,707 entries with 28 features.
2. A preview of the dataset information, as well as null value check revealed a few things for us to look further:
    a. Date / time features (including 'Date Rptd', 'DATE OCC', 'TIME OCC') need to be fixed to be datetime type before being used as a time dimension. Further, 'TIME OCC' may need more care, given it should be in military hours but is currently in 'int' format.
    b. Be aware that the numbers in the features with 'code' in nature (such as 'Crm Cd', 'Premis Cd', 'Weapon Used Cd') only represent certain types of that feature, although they are of 'int' or 'float' data type currently, no operation should be applied.
    c. There is a decent amount of missing value for descriptive features, led by additional crime types (e.g. 'Crm Cd 4', 'Crm Cd3', 'Crm Cd2'), followed by location, weapon and victim descriptions. Luckily we may not necessarily use those features. But if we do use any of them, we will drop the entries with NaN before conducting any analysis.
    d. There are features which have 'object' in Dtype. We believe it's due to mixed type values, such as mixture of NaN and strings (example: 'Cross Street')

Table 1. General information of dataset for EDA purpose

Primary information of the dataset

```
The primary information of the dataset:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 910707 entries, 0 to 910706
Data columns (total 28 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   DR_NO           910707 non-null  int64
 1   Date Rptd       910707 non-null  object
 2   DATE OCC        910707 non-null  object
 3   TIME OCC        910707 non-null  int64
 4   AREA            910707 non-null  int64
 5   AREA NAME       910707 non-null  object
 6   Rpt Dist No     910707 non-null  int64
 7   Part 1-2        910707 non-null  int64
 8   Crm Cd          910707 non-null  int64
 9   Crm Cd Desc     910707 non-null  object
 10  Mocodes         783696 non-null  object
 11  Vict Age        910707 non-null  int64
 12  Vict Sex        789672 non-null  object
 13  Vict Descent    789663 non-null  object
 14  Premis Cd       910697 non-null  float64
 15  Premis Desc     910153 non-null  object
 16  Weapon Used Cd  315247 non-null  float64
 17  Weapon Desc     315247 non-null  object
 18  Status          910707 non-null  object
 19  Status Desc     910707 non-null  object
 20  Crm Cd 1        910696 non-null  float64
 21  Crm Cd 2        66335 non-null   float64
 22  Crm Cd 3        2237 non-null    float64
 23  Crm Cd 4        64 non-null      float64
 24  LOCATION        910707 non-null  object
 25  Cross Street    143332 non-null  object
 26  LAT             910707 non-null  float64
 27  LON             910707 non-null  float64
dtypes: float64(8), int64(7), object(13)
memory usage: 194.5+ MB
```

Null value summary by features

```
Null value summary:
Crm Cd 4           910643
Crm Cd 3           908470
Crm Cd 2           844372
Cross Street       767375
Weapon Used Cd     595460
Weapon Desc        595460
Mocodes            127011
Vict Descent       121044
Vict Sex           121035
Premis Desc           554
Crm Cd 1               11
Premis Cd              10
LAT                     0
LOCATION                0
Status Desc             0
Status                  0
DR_NO                   0
Date Rptd               0
Vict Age                0
Crm Cd Desc             0
Crm Cd                  0
Part 1-2                0
Rpt Dist No             0
AREA NAME               0
AREA                    0
TIME OCC                0
DATE OCC                0
LON                     0
dtype: int64
```

3. For our purpose, the most useful features / columns should be:
    - Report number (DR_NO), Date occurred (DATE OCC), date reported (Date Rptd), time occurred (TIME OCC), area name (AREA NAME), crime description (Crm Cd Desc), victim age (Vict Age), victim sex (Vict Sex), victim descendance (Vict Descent).
    - Latitude & Longitude columns are useful in mapping the crime occurrence and density in the map of LA
4. A feature-by-feature 'value_counts' examination revealed below key observations:
    a. All entries can be uniquely identifiable by division record number ('DR_NO')

b. 'Victim age' has some potential invalid entries, such as those with values of '0, -2, -3, or -4), which need to be removed before using this feature.
c. 'Vict Sex' and 'Vict Descent' contain unidentifiable values '-', which need to be removed before using any of them.

Table 2. The features with correction needs revealed by 'value_counts' examination

DR_NO value counts	Vict Age value counts	Vict Sex value counts	Vict Descent value counts
DR_NO 190326475 1 221908151 1 231105297 1 221008844 1 200320258 1 .. 240405919 1 240905054 1 241605414 1 241605270 1 240104953 1 Name: count, Length: 910707,	Vict Age 0 229674 30 20613 35 20236 31 19705 29 19667 ... 97 67 -2 17 -4 3 -3 2 120 1	Vict Sex M 373417 F 332804 X 83347 H 103 - 1	Vict Descent H 277284 W 184107 B 127950 X 91720 O 71864 A 19994 K 4965 F 3910 C 3623 J 1279 V 967 I 867 Z 456 P 243 U 183 D 73 G 65 L 62 S 49 - 2

Data Cleanup:

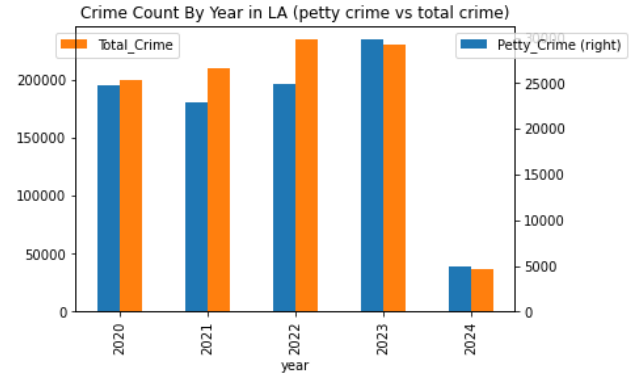
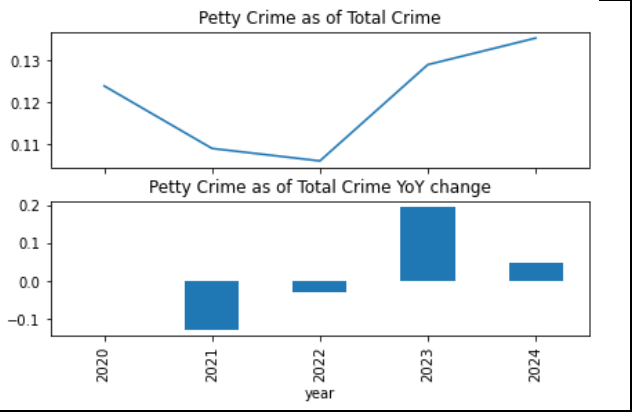
- Change the type of 'DATE OCC' and 'Date Rptd' to datetime format.
- First change the type of 'TIME OCC' from int to string with leading zeros with four digits length (so that it becomes 24-hour military time per documentation), then change each entry to pandas time format.
- Drop all other columns that are not going to be used for future analysis to keep the dataset size nimble. We choose to keep the columns as below:
  - ['DR_NO', 'DATE OCC', 'Date Rptd', 'TIME OCC', 'AREA NAME', 'Premis Desc', 'Crm Cd Desc', 'Vict Age', 'Vict Sex', 'Vict Descent']
- For the invalid values existing in 'Vict Age', 'Vict Sex', 'Vict Descent', we choose to only drop them when we are actively using those features for analysis, considering dropping them may reduce the sample size meaningfully (for example, age with '0' as value has ~23k entries)
- After the fixing, our dataset reduces to 910,707 entries (the same as before) and 9 columns (from 28).

Descriptive Hypothesis Testing:

- **Hypothesis One** - Petty crimes in Los Angeles, defined as those with an economic loss of less than $950, increased meaningfully in recent years.
  - **Background**: Anecdotally, we heard that LA petty crimes increased meaningfully after 2020, is there evidence to support this hypothesis?
    - Reference: [California Penal Code § 484(a) PC – Petty Theft]
  - **Analysis**: In this case, we would like to use the dataset on hand to examine the statement descriptively. Specifically, we compare three related concepts across time: total petty crime count, total crime count, and the proportion of petty crime as percentage of total crime count.

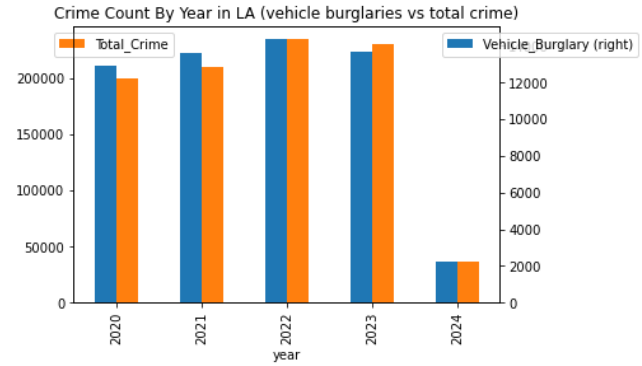
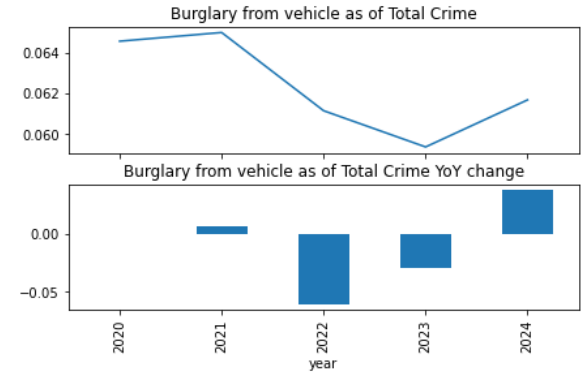
- **Finding**: first, we found that both total crime and petty crime yearly count increased since 2020. However, more distinctly, petty crime count picked up more meaningfully in 2023, while total crime count stayed about the same. As a result, the proportion of petty crime as percentage of total crime notably increased in 2023, after dipping in 2021 first.
- **Conclusion:** From observation: data seems to support the argument that petty crime in LA has picked up in recent years, but more specifically, it has picked up as a proportion of total crime in 2023.

Fig 1. Key observations for Hypothesis One testing

Crime Count By Year – Petty vs Total Crime	Petty Crime as % of Total Crime Per Year
	

- **Hypothesis Two** - Burglaries from vehicles increased meaningfully in recent years.
  - **Background**: Anecdotally, we heard that burglaries from vehicles increased meaningfully across California in recent years, is there evidence to support this hypothesis?
  - **Analysis**: Similarly, we would like to use the LA dataset to examine the statement descriptively. Specifically, we compared the concepts below across time: total burglaries from vehicles count, total crime count, and the proportion of burglaries from vehicles as percentage of total crime count.
  - **Findings**: We found that by observation, vehicle burglaries count stayed relatively stable with small increases in the past three years. However, as total crime counts increased, the proportion of vehicle burglaries decreased slightly in 2022. Although such a percentage increased slightly in 2024, given there are only two months' data, we would not draw much conclusion from it.
  - **Conclusion:** From observation: data doesn't seem to support the argument that burglaries from vehicles in LA picked up in most recent years, neither from total counts, nor as a percentage of the total crime.

Fig 2. Key observations for Hypothesis Two testing

Crime Count By Year – Vehicle Burglaries vs Total	Vehicle Burglaries as % of Total Crime
	

Outcome Based Questions:

1. **Is there any pattern associated with the time of the day, the day of a month, or the Month of a year?**

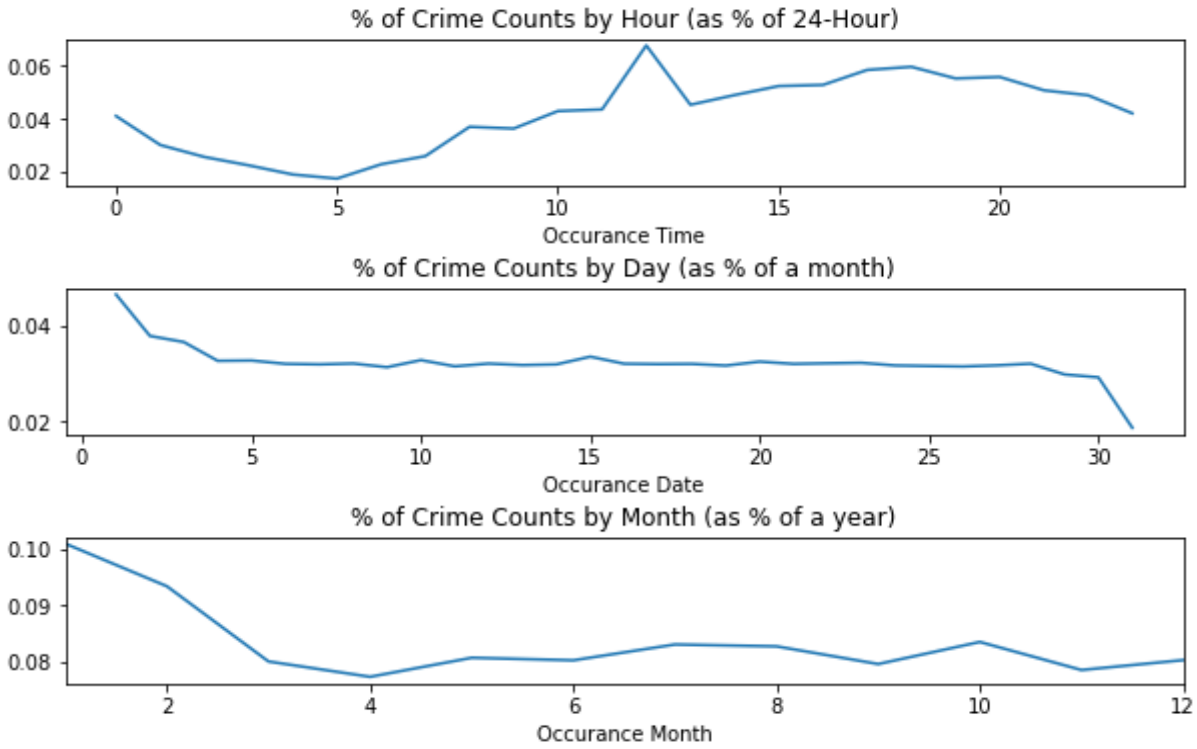
   Analysis: We utilized the crime occurrence time, and occurrence day, group the crime counts by the time of the day, the day of a month, and the month of a year, then compute the percentage of the sub-time window's crime count as a percentage of the larger time window (e.g. % of crime count by hour in the 24-hour window)

   Findings: We noted that *[Fig 3]*

   i. Late afternoon and evening tend to have the highest crime rate of a day.
   ii. The first five days of the month tend to have a higher crime rate while the month end tends to have a lower crime rate.
   iii. January - February has a notably higher rate than the rest of the year.

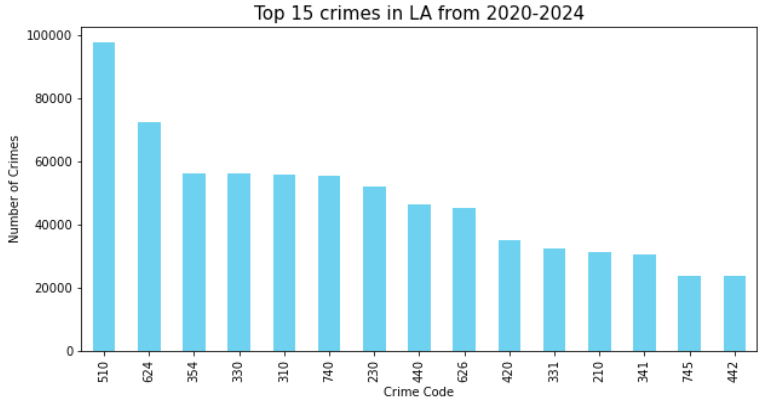
   Implications: help agencies to effectively allocate resources for periodic surges in crimes.

Fig 3. Key observations for time related pattern for LA crime occurrence



2. **What are the major incidents of crime in LA?**
   Analysis: Ranking the highest 15 crimes by count since 2020, and calculated their contribution to the % of total crimes in LA.
   Findings:
   1. Highest 15 Crime Codes make up about 78.3% of all crimes in LA. (Top 10 - ~62%.).
   2. Highest crime in LA is "VEHICLE - STOLEN," Code 510 and contributes to 10.73% of all crimes between 2020 Jan - 2024 Mar. *[Fig 4]*

Fig 4 - Highest 15 Crimes in LA



Crm Cd	Crm Cd Desc	Crime Count
510	VEHICLE - STOLEN	97716
624	BATTERY - SIMPLE ASSAULT	72353
354	THEFT OF IDENTITY	56254
330	BURGLARY FROM VEHICLE	56134
310	BURGLARY	55862
740	VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS)	55219
230	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	51788
440	THEFT PLAIN - PETTY ($950 & UNDER)	46432
626	INTIMATE PARTNER - SIMPLE ASSAULT	45277
420	THEFT FROM MOTOR VEHICLE - PETTY ($950 & UNDER)	35036
331	THEFT FROM MOTOR VEHICLE - GRAND ($950.01 AND OVER)	32526
210	ROBBERY	31122
341	THEFT-GRAND ($950.01 & OVER)EXCPT,GUNS,FOWL,LIVESTK,PROD	30329
745	VANDALISM - MISDEAMEANOR ($399 OR UNDER)	23648
442	SHOPLIFTING - PETTY THEFT ($950 & UNDER)	23516

3. **What has been the YoY trend in crime reported in each LAPD Division?**

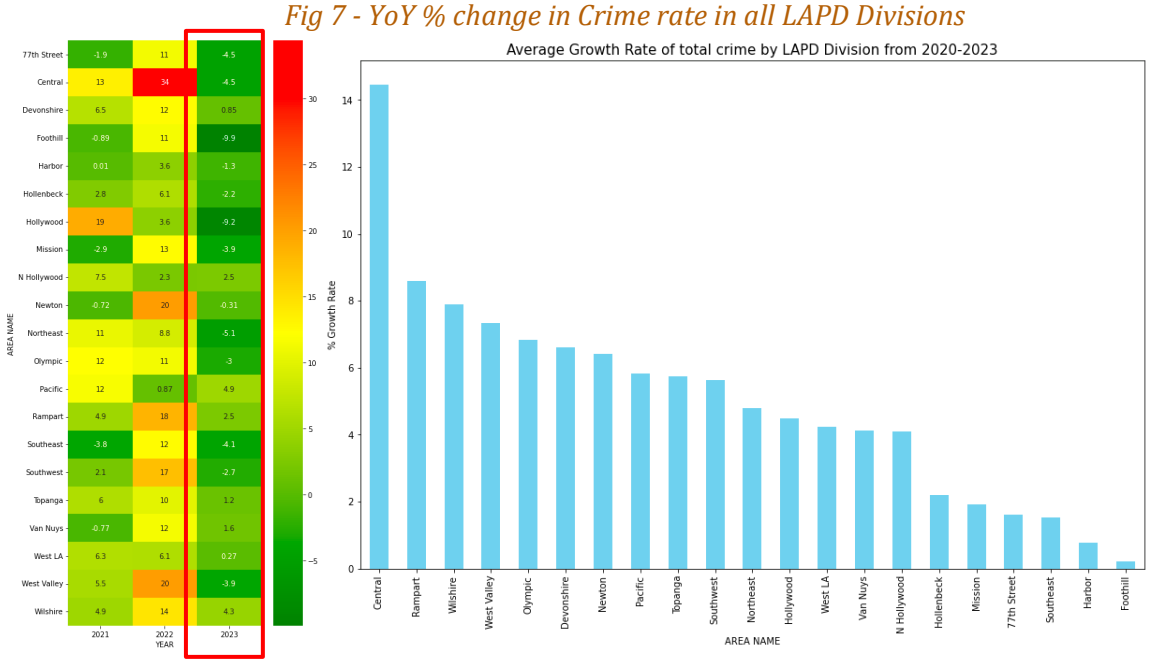
   Analysis: Listed the Total Crimes by area between 2020-2023 and the % YoY change

   Findings:

   i. Between 2021-2022, Central LAPD division saw a 34.4% spike, while Newton & West Valley saw a 20% increase in crime. *[Fig 7]*

   ii. Crime has reduced in all LAPD Divisions in 2023, except Devonshire, N-Hollywood, Pacific, Rampart, Topanga, VanNuys & Wilshire.

Fig 7 - YoY % change in Crime rate in all LAPD Divisions



4. **Is there any relationship between the type of crime and districts covered by LAPD Divisions?**

   a. **What is the highest crime recorded in each LAPD Division/Area?**

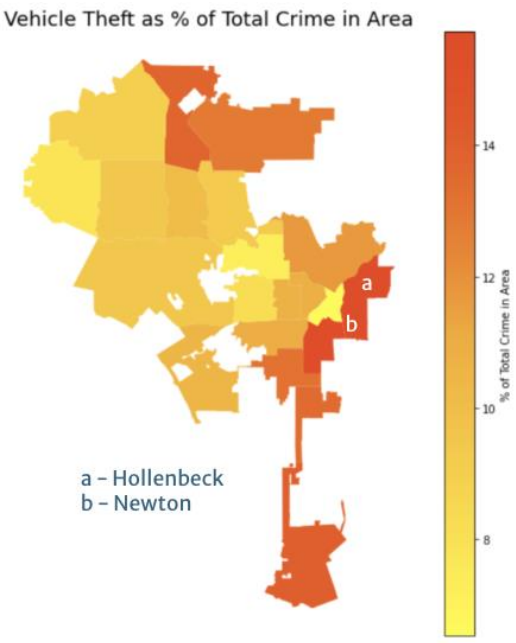
   Analysis: Mapped the highest 5 crimes by crime code/description in each area and calculated the % contribution to all crimes reported in that area.

   Findings:

   i. VEHICLE - STOLEN is the highest crime averaging >10% in 14/20 Divisions, reaching as high as 15% in Hollenbeck & Newton. *(Fig 5)*

   ii. BURGLARY & BURGLARY FROM VEHICLE is the highest in 5 of the 20 LAPD Divisions.

Fig 5 - Vehicle Thefts report in LAPD Divisions as % of total crime in that area



AREA NAME	Vehicle Theft Count	Area Total Crime	Percentage of Vehicle Theft
NEWTON	7151	45475	15.73
HOLLENBECK	5317	34021	15.63
HARBOR	5248	37470	14.01
MISSION	4985	36201	13.77
SOUTHEAST	6169	45944	13.43
77TH STREET	7488	56887	13.16
FOOTHILL	3864	30123	12.83
NORTHEAST	4608	39236	11.74
SOUTHWEST	5612	51174	10.97
RAMPART	4646	42981	10.81
OLYMPIC	4955	46005	10.77
PACIFIC	5588	53098	10.52
VAN NUYS	3876	38445	10.08
WEST VALLEY	3638	38241	9.51
N HOLLYWOOD	4257	45618	9.33
DEVONSHIRE	3339	37172	8.98
WILSHIRE	3583	43676	8.20
TOPANGA	2888	37176	7.77
HOLLYWOOD	3448	47934	7.19
WEST LA	3002	41759	7.19
CENTRAL	4054	62071	6.53

b. **What has been the YoY trend in vehicle thefts between 2020-2023?**

Analysis: Summarizing average growth rate in stolen vehicles (Cd 510) YoY by area

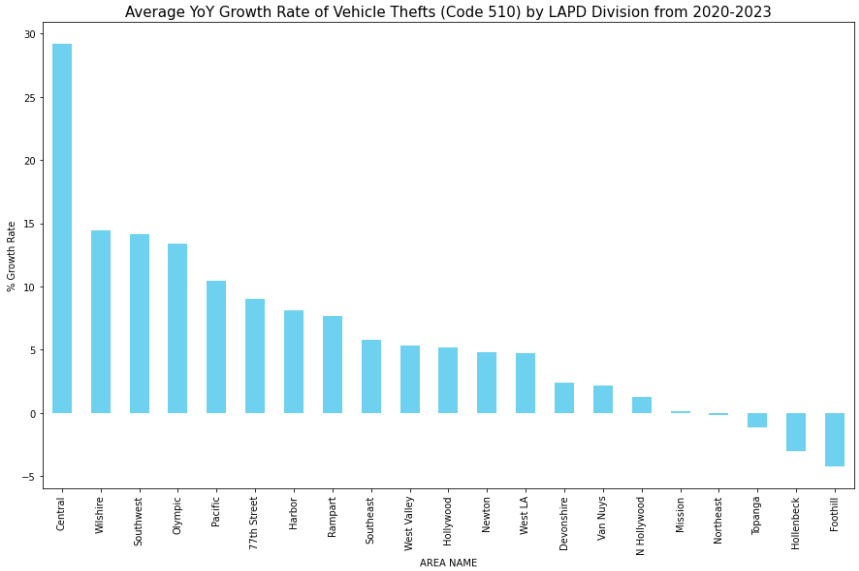
Findings (Map Reference Source): *[Fig 8]*

i. Central has the highest YoY growth rate in Vehicle Thefts - ~30%.

ii. Wilshire, Southwest & Olympic have seen 14-15% avg rise in Vehicle Thefts.

iii. Only Foothill, Hollenbeck & Topanga have shown decline in vehicle thefts..

Fig 8 - Avg YoY growth rate in Vehicle Thefts in each LAPD Division



5. **Is there any relationship between crimes with firearms and LAPD Divisions/Area?**

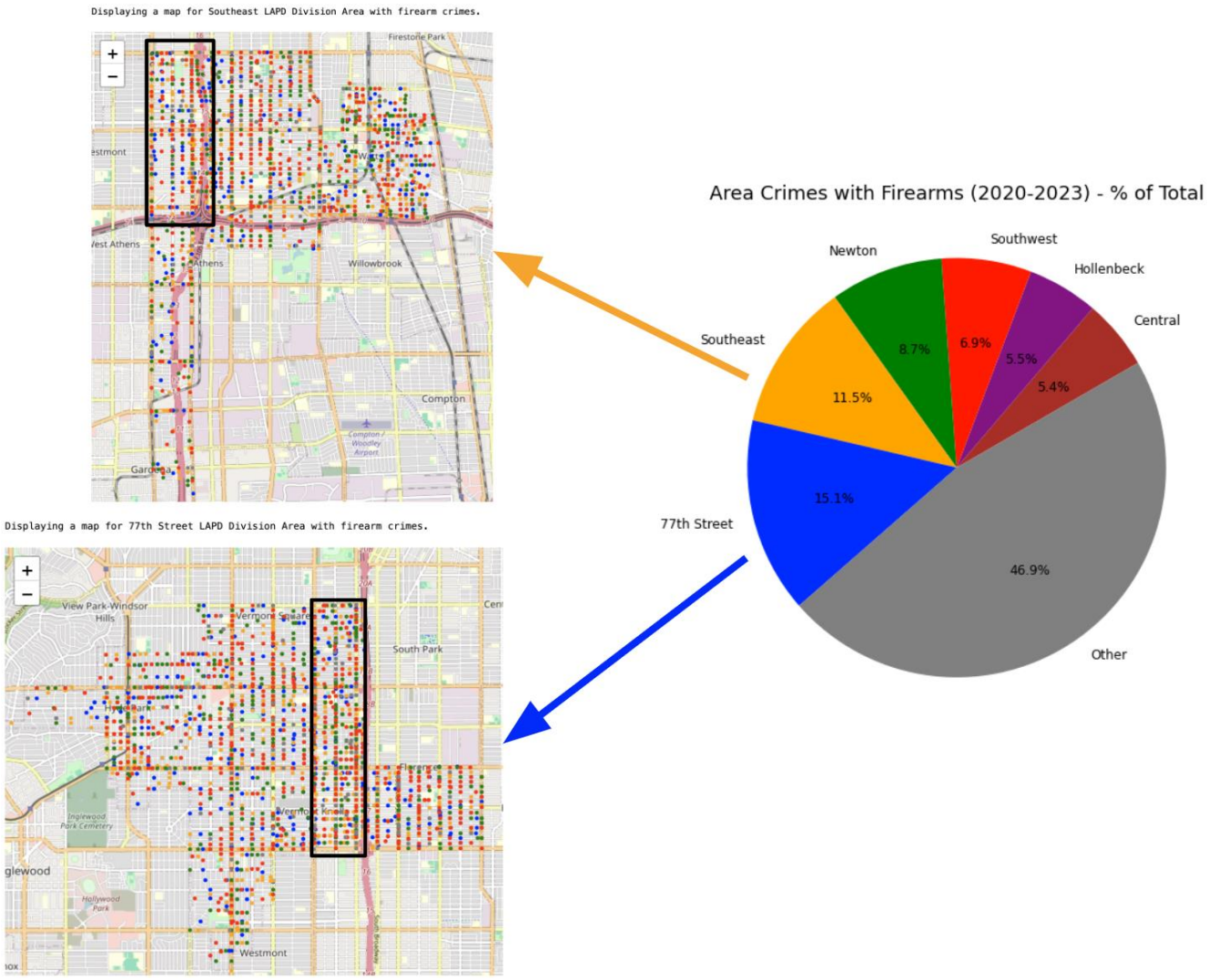
Analysis: Summarized crimes with firearms in each area and as a % of total firearm crimes in LA from 2020-2023.

Findings:

i. The highest 6 LAPD Divisions recorded 53% of all firearm crimes in LA.
ii. 77th Street and Southeast LAPD have the highest occurrences and make up 26.49% of crimes with firearms in LA.
iii. The 5 blocks grid *(area outlined in map)*, between S Vermont Ave, Harbor Freeway and Century Freeway has the highest density of crimes with firearms in the 2 Areas covered by 77st St & Southeast LAPD divisions.

Fig 6 - Crime Count % w/ firearms by area

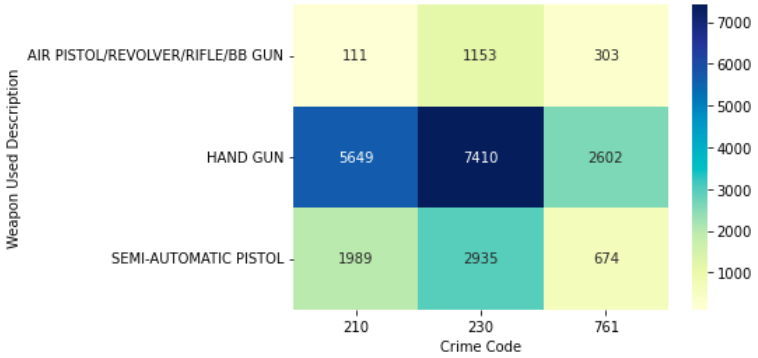
{2020: 'blue', 2021: 'orange', 2022: 'green', 2023: 'red', 2024: 'grey'}



6. **Is there a relationship between the type of crime and the weapon used?**
Analysis: We listed the highest 3 known firearms and the highest 3 crimes committed with them.
Findings: *[Fig 9]*

i. Highest count of crimes with a firearm - Crime code 230 - Assault with a deadly weapon/Aggravated Assault. Followed by Robbery (Crime Code 210).
ii. The most commonly used weapon is a handgun followed by a semi-automatic pistol [Fig 9.1]

Fig 9 - Highest 3 known firearms & crimes



Crm Cd Desc	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	BRANDISH WEAPON	ROBBERY
Weapon Desc			
AIR PISTOL/REVOLVER/RIFLE/BB GUN	1153	303	111
HAND GUN	7410	2602	5649
SEMI-AUTOMATIC PISTOL	2935	674	1989

Limitations:

We understand that this is primarily a descriptive / exploratory analysis, which does not contain statistical testing and inference by design. Therefore it can serve as a starting point for any further statistical analysis and research for parties who are interested.

Key takeaways:

This study explores LA crime log from 2020 to March, 2024 to assess if the data supports the public perceptions regarding the crime trend in LA. It also contains descriptive analysis that highlights trends, patterns & insights that could be of help to local law-enforcement.

- Data supports the argument that petty crime in LA has increased in recent years, especially in 2023. However, it does not support the argument that burglaries from vehicles have risen in the same timeframe.
- For time associated crime patterns, late afternoon and evening, the first five days of the month, and the months of January to February have a higher crime rate, in a day/month/year respectively. While the month end tends to have a lower crime rate.
- Vehicle Theft (Crm Cd 510) is the highest crime in LA. 16% of all crimes reported in Newton & Hollenbeck LAPD divisions is a Stolen Vehicle.
- Between 2022-2023 total crime has reduced in all LAPD Divisions except these 7 divisions - Devonshire, N-Hollywood, Pacific, Rampart, Topanga, VanNuys & Wilshire.
- 6 LAPD Divisions recorded 53% of all firearm crimes in LA.
    - topped by 77st St & Southeast LAPD divisions (26% of all firearm crimes in LA).