Analyzing the relationship between age, sex, and identity theft reports in the city of Los Angeles, California (2023)

Ashley Klibowitz
November 2024

Executive Summary

Introduction
This analysis examines the relationship between age group, sex, and identity theft reports in Los Angeles in 2023. It highlights the highest and lowest report volumes by demographic and compares these with chi square test expectations.

Background and Research Question
The research question: Is there a relationship between age group, sex, and the number of identity theft reports in Los Angeles in 2023? is explored using a dataset provided by the Los Angeles Police Department.

Data Collection and Structure
The dataset includes 28 columns. Key variables for analysis include date, victim age, victim sex, and crime description.

Data Cleaning
Data cleaning involved filtering for 2023 identity theft cases, correcting "Vict Sex" inconsistencies, and removing invalid or missing age data, resulting in 10,966 valid records.

Data Analysis
Descriptive statistics showed a right-skewed distribution, with most reports from younger demographics. The chi square test found a significant relationship between age group, sex, and identity theft reports, and revealed substantial differences from expected values in the test.

Discussion and Conclusions
The analysis shows that demographic factors impact identity theft report volumes in Los Angeles. External studies reveal that females are more likely to report identity theft, which aligns with the finding that they have the highest number of reports. Additionally, individuals in their 30s are the most affected by identity theft.

Introduction
This analysis examines the relationship between age group, sex, and the number of identity theft reports filed in Los Angeles, California, in 2023. By focusing on report counts, it identifies expected patterns for each group and uncovers potential disparities, providing a foundation for targeted public awareness initiatives.

The increasing prevalence of identity theft prompted this investigation. Los Angeles, California, with its diverse demographics, presents a unique opportunity to examine the relationship between age group, sex, and number of identity theft reports in 2023.

Data Collection
The data in the dataset "Los Angeles Crime Data from 2020 to Present" was collected from the Los Angeles Police Department (LAPD). The dataset includes crime incidents from 2020 to the present recorded and updated by LAPD.

The dataset can be accessed here: [Crime Data from 2020 to Present](#).

Data Structures
The dataset is structured in a tabular CSV file format with 28 columns representing various attributes of crime incidents. Key columns for this analysis include:

- **Vict Age (Text):** Two-character numeric age of the victim; 0 indicates unknown.
- **Vict Sex (Text):** Victim's sex as F (Female), M (Male), or X (Unknown).
- **Crm Cd Desc (Text):** Crime description, e.g., "THEFT OF IDENTITY."
- **DATE OCC (Floating Timestamp):** Date the crime occurred (MM/DD/YYYY).

The full list of columns and their data types includes:
- **DR_NO (Text):** Official file number.
- **Date Rptd (Floating Timestamp):** Date reported (MM/DD/YYYY).
- **DATE OCC (Floating Timestamp):** Date occurred (MM/DD/YYYY).
- **TIME OCC (Text):** Time in 24-hour format.
- **AREA (Text):** LAPD Community Police Station number.
- **AREA NAME (Text):** Police area name.
- **Rpt Dist No (Text):** Sub-area code.
- **Part 1-2 (Number):** Crime classification.
- **Crm Cd (Text):** Crime code.
- **Crm Cd Desc (Text):** Crime description.
- **Mocodes (Text):** Suspect modus operandi.
- **Vict Age (Text):** Victim's age.
- **Vict Sex (Text):** Victim's sex.
- **Vict Descent (Text):** Victim descent code.
- **Premis Cd (Number):** Location type.
- **Premis Desc (Text):** Premise description.
- **Weapon Used Cd (Text):** Weapon type.

- **Weapon Desc (Text):** Weapon description.
- **Status (Text):** Case status.
- **Status Desc (Text):** Case status description.
- **Crm Cd 1 (Text):** Primary crime code.
- **Crm Cd 2-4 (Text):** Additional crime codes.
- **LOCATION (Text):** Incident address.
- **Cross Street (Text):** Cross street.
- **LAT (Number):** Latitude.
- **LON (Number):** Longitude.

Data Cleaning

To clean the dataset, the following steps were taken in OpenRefine:

1. The "DATE OCC" column was formatted as text, so it was transformed into a date format using common transforms.
2. The timeline facet was used to filter entries from 2023, which were starred. Non-starred entries were deleted, leaving only relevant records.
3. The "Crm Cd Desc" field was filtered with the text facet to isolate identity theft-related records, which were starred. Non-relevant entries were deleted to focus on identity theft cases.
4. Invalid entries in the "Vict Sex" column, such as "H," blank values, or "X," were identified utilizing the test facet function and removed to maintain data accuracy.
5. The "Vict Age" column contained negative values (e.g., -1, -2) and "0" as placeholders for missing data. These were deleted to ensure only valid age entries remained.
6. Irrelevant columns were removed, leaving only "Crm Cd Dec," "Vict Age," and "Vict Sex" for analysis. The "DATE OCC" column was excluded as it was already used to obtain only 2023 entries.

Data Analysis

Descriptive Statistics
The cleaned dataset, comprising 10,966 entries, was imported into Excel. Descriptive statistics were calculated to guide the selection of an appropriate test, which was determined to be the chi square test.

Descriptive Statistics Calculations		
Mean (Average)		
42.08035		
Median		
38		
Mode		
32		
Standard Deviation		
14.99035		
Varience		
224.7105		
Range		
79		
Skewness		
0.886753		
Kurtosis		
0.159885		

Figure 1 shows the descriptive statistics calculations.

The data showed a right-skewed distribution (positive skewness of 0.89), with most reports in younger age groups. The mean age (42.08) exceeded the median (38), reflecting the influence of larger data points. The kurtosis suggested a higher peak and heavier tail than a normal distribution. The histogram confirmed the right-skewed pattern.
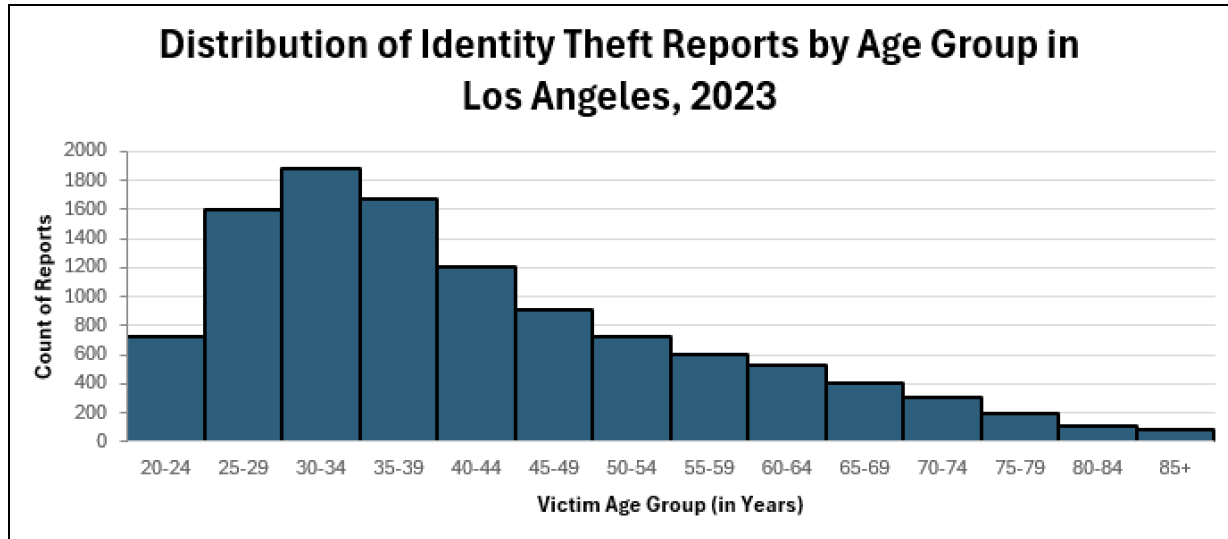
Figure 2 shows the age distribution of identity theft victim reports.

Chi Square Test
A chi square test was performed to examine the relationship between age, sex, and identity theft victimization. The chi square test was chosen for this analysis because it's a non-parametric method suitable for categorical data.

The age groups were evenly categorized in 4-year spans to ensure sufficient sample sizes for reliable analysis. Larger ranges, like 85+, were used for older populations with fewer individuals.

Hypotheses:
- **Null Hypothesis (H_0):** There's no relationship between age, sex, and the number of identity theft reports in Los Angeles, California, 2023.
- **Alternative Hypothesis (H_1):** There's a relationship between age, sex, and the number of identity theft reports in Los Angeles, California, 2023.

Observed Values

Age Group	Male Victims	Female Victims	
20-24	234	490	**724**
25-29	514	1080	**1594**
30-34	609	1277	**1886**
35-39	647	1024	**1671**
40-44	513	693	**1206**
45-49	388	524	**912**
50-54	333	392	**725**
55-59	308	295	**603**
60-64	285	240	**525**
65-69	190	220	**410**
70-74	148	162	**310**
75-79	88	113	**201**
80-84	51	56	**107**
85+	27	64	**91**
	4335	**6630**	**10965**

Figure 3 represents the actual data collected from the study.

Expected Values

Age Group	Male Victims	Female Victims	
20-24	286.2326	437.7674	**724**
25-29	630.186	963.814	**1594**
30-34	745.6279	1140.372	**1886**
35-39	660.6279	1010.372	**1671**
40-44	476.7907	729.2093	**1206**
45-49	360.5581	551.4419	**912**
50-54	286.6279	438.3721	**725**
55-59	238.3953	364.6047	**603**
60-64	207.5581	317.4419	**525**
65-69	162.093	247.907	**410**
70-74	122.5581	187.4419	**310**
75-79	79.46512	121.5349	**201**
80-84	42.30233	64.69767	**107**
85+	35.97674	55.02326	**91**
	4335	**6630**	**10965**

Figure 4 values reflect if there was no relationship between the variables.

Chi Square Calculation ((actual - expected) ^2 / expected)

Age Group	Male Victims	Female Victims
20-24	9.531551	6.232168
25-29	21.42097	14.00602
30-34	25.03552	16.36938
35-39	0.281126	0.183813
40-44	2.749872	1.797993
45-49	2.088583	1.365612
50-54	7.502309	4.905356
55-59	20.32258	13.28784
60-64	28.89427	18.89241
65-69	4.804644	3.141498
70-74	5.281479	3.453275
75-79	0.916682	0.599369
80-84	1.788307	1.169278
85+	2.239834	1.464507

Figure 5 values are derived from the chi square calculation using both observed and expected values.

219.7262	chi square value to test against critical value.
1.19E-39	calculated p-value, this is much lower than 0.05 so significant.
1.19E-39	calculated p-value using CHITEST.
219.7262	chi square value from the chi square table using calculated p-value.
22.36203	critical value from the chi square table using significance level of 0.05.
	chi-square value exceeds this critical value.
	null hypothesis is rejected.

Figure 6 shows the results of the chi square calculations.

The chi square value at 219.7262 indicates the observed values significantly deviate from the expected values. Since the chi square value exceeds the critical value of 22.362, the null hypothesis is rejected, and the alternative hypothesis is accepted that there's a significant relationship between age, sex, and identity theft victimization.
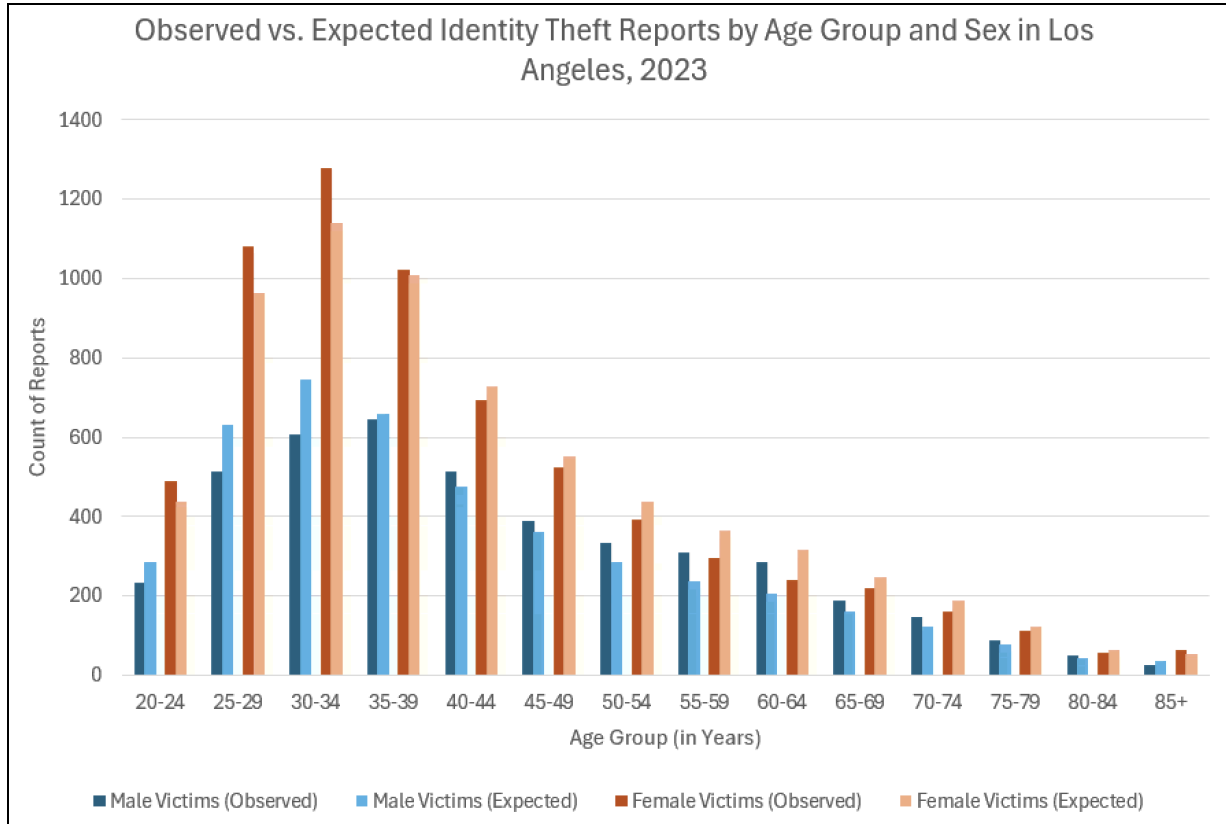
Figure 7 displays a comparison of the observed and expected identity theft reports for both males and females, categorized by age group.

Female victims aged 25-34 significantly exceed expected values, while male victims in the same age groups significantly fall below expectations. The 35-39 age group aligns the closest with expected values, showing minimal differences for both males and females.
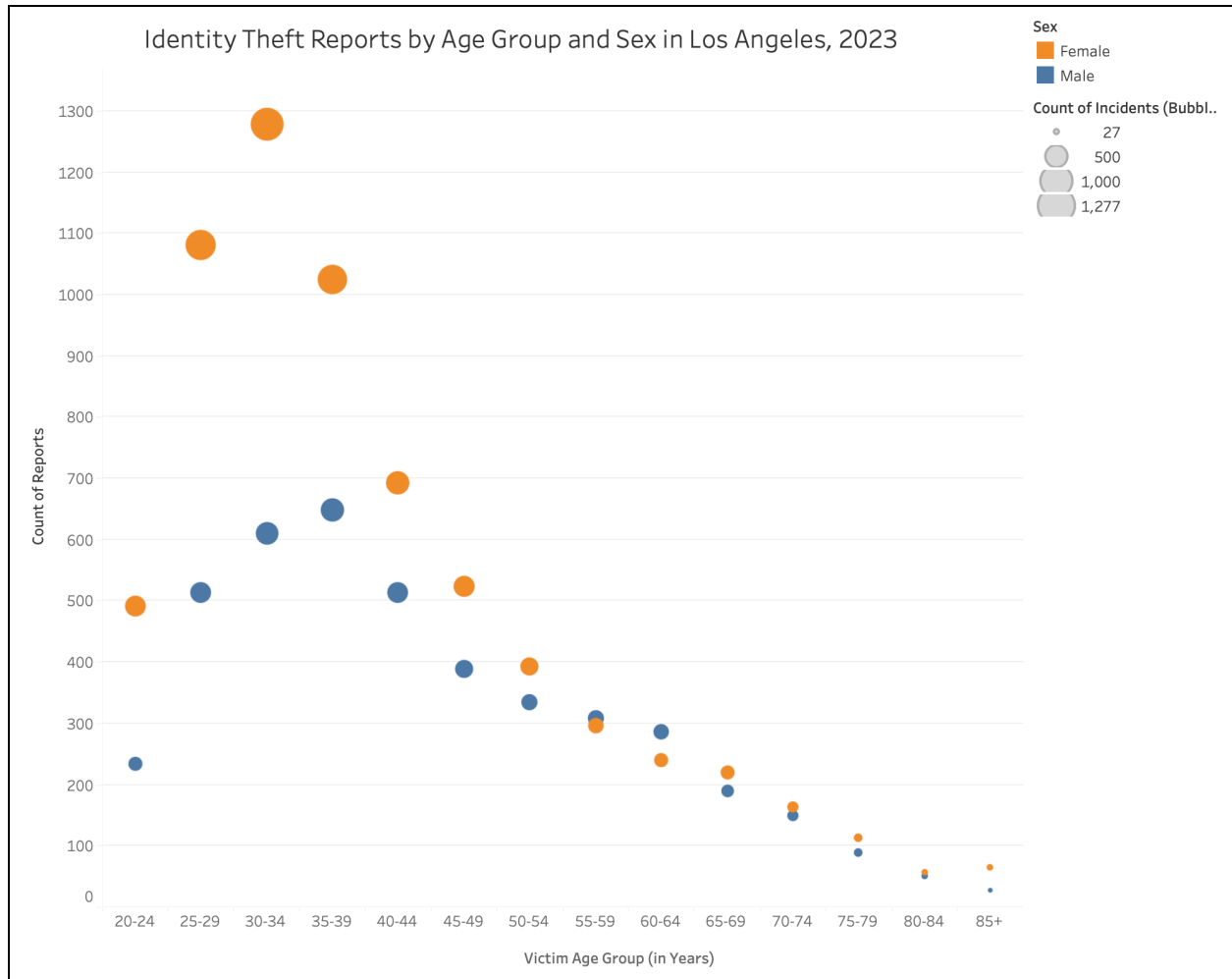
Figure 8 shows a bubble chart illustrating the relationship between age, sex, and the actual number of identity theft reports in Los Angeles, California, in 2023.

The bubble chart shows that identity theft reporting patterns shift with age and sex. While females predominantly report higher numbers in younger age groups, males surpass them in some older demographics.

The chart shows that females report the most identity theft cases in the 30-34 age group, with 1,277 cases, and outnumber males in 12 out of 14 age groups. Males report their highest total in the 35-39 age group, with 647 cases, and in the 55-59 and 60-64 age groups report more cases than females. Additionally, females and males are the most aligned at 55-59. This demonstrates how identity theft reporting differs across age groups and sex.

Discussion and Conclusions

In conclusion, this analysis examined the relationship between age, sex, and identity theft in Los Angeles in 2023. Descriptive statistics showed a right-skewed distribution, with the highest concentration of victims in the 30-34 age group. The chi square test confirmed that age and sex significantly influenced identity theft reports, with females in the 25-34 age group significantly exceeding expected values and males falling below.

U.S. Census data shows an equal male-to-female ratio in Los Angeles, suggesting other factors beyond population size may be influencing the reports (U.S. Census Bureau, 2023). The Australian Communications Consumer Action Network (ACCC) finds that females are 50% more likely to report identity theft (Australian Communications Consumer Action Network, 2016). These patterns may apply to Los Angeles, suggesting that females may be more inclined to report identity theft.

A report by The Federal Trade Commission shows that individuals aged 30-39 are the most affected by identity theft (Federal Trade Commission, 2023). This matches the analysis showing the highest number of total reports in this range. Future analysis might include examining behavioral, societal, or situational factors that could drive these trends.

Exploring additional variables, such as socioeconomic status, education, and lifestyle, could provide a deeper understanding of the overall patterns observed. Insights into the type of identity theft (e.g., social media fraud, phishing) would also be useful, but the dataset was limited. Despite these limitations, the findings in this analysis still offer valuable insights for targeted public awareness initiatives.

References

Australian Communications Consumer Action Network. (2016). Identity theft and gender
[Infographic].
https://accan.org.au/files/Grants/ANU%20ID%20theft/ANU%20ID%20theft%20infographic_Gender.pdf

City of Los Angeles. (2024). Crime data from 2020 to present [Data set]. Data.gov.
https://catalog.data.gov/dataset/crime-data-from-2020-to-present

Federal Trade Commission. (2023). Consumer Sentinel Network data book 2023 [PDF]. Federal
Trade Commission.
https://www.ftc.gov/system/files/ftc_gov/pdf/CSN-Annual-Data-Book-2023.pdf

U.S. Census Bureau. (2023). QuickFacts: Los Angeles city, California. U.S. Department of
Commerce. Retrieved November 16, 2024, from
https://www.census.gov/quickfacts/losangelescitycalifornia