



INSTITUTO FEDERAL DE SÃO PAULO
TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

ANÁLISE EXPLORATÓRIA DE DADOS
Turma de Quinta-feira

Giovanna Camille Silva Carvalho SP3123162
Kelly Radchelle Araujo de Souza SP3123588

São Paulo
2025

Sumário

1. Introdução ao Tema.....	3
2. Descrição Dos Dados.....	3
3. Metodologia.....	3
4. Análises e Resultados.....	4
4.1. Perfil e Distribuição dos Criminosos Sexuais Registrados.....	4
4.1.1. Quantidade de criminosos sexuais registrados por Community Area....	4
4.1.2. Distribuição etária dos criminosos sexuais com e sem vítimas menores.....	6
4.1.3. Perfil demográfico (raça, gênero) dos infratores sexuais com e sem vítimas menores.....	8
4.1.4. Faixas etárias mais associadas a vítimas menores e distribuição geográfica.....	10
4.1.5. Crimes prevalentes e perfis demográficos nas áreas com mais criminosos sexuais.....	12
4.2. Análise Geográfica e Comparação Entre Áreas.....	15
4.2.1. Community Areas com alta frequência de crimes gerais e alta concentração de criminosos sexuais.....	15
4.2.2. Diferença na distribuição dos tipos de crimes entre áreas com alta e baixa concentração de criminosos sexuais.....	16
4.2.3. Comparação dos três principais crimes em áreas de alta e baixa incidência de crimes sexuais.....	18
4.3. Análises Temporais.....	20
4.3.1. Relação entre a tendência dos crimes sexuais e crimes gerais ao longo dos anos.....	20
4.3.2. Variação anual da proporção de crimes sexuais sobre o total de crimes. 22	
4.4. Testes de Hipótese e Inferência Estatística.....	26
4.4.1. Diferença na média anual de crimes entre áreas com alta e baixa concentração de criminosos sexuais.....	26
4.4.2. Probabilidade de ocorrência de crimes sexuais em áreas com mais de um criminoso registrado.....	28
4.4.3. Probabilidade de um crime ocorrer em área com pelo menos um criminoso sexual registrado.....	29
5. Conclusão.....	31
6. Referências.....	32

1. Introdução ao Tema

Este documento apresenta uma análise exploratória de dados sobre crimes na cidade de Chicago, com ênfase em crimes sexuais registrados a partir de 2011. Os dados foram obtidos no portal <https://catalog.data.gov/dataset/>, base oficial de dados do governo dos Estados Unidos. O objetivo é aplicar princípios de probabilidade e estatística para investigar se existem regiões com simultânea alta criminalidade geral e presença significativa de infratores sexuais registrados, gerando assim, referências para políticas públicas de segurança.

2. Descrição Dos Dados

Os dados analisados abrangem registros de crimes em geral e, especificamente, de infrações sexuais na cidade de Chicago. O conjunto inclui variáveis de localização, data e hora do incidente, além de informações pessoais do suspeito, como gênero, raça e data de nascimento. As bases completas podem ser consultadas nos links abaixo:

- Crimes (2001-presente):
https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2/about_data
- Sex Offenders:
https://data.cityofchicago.org/Public-Safety/Sex-Offenders/vc9r-bqvy/data_preview

A fim de conseguir resultados estatísticos mais precisos, foi utilizado um conjunto da base “Crimes (2001-presente)”, cujos dados foram registrados somente entre os anos de 2011 a 2025, mesmo período em que a base “Sex Offenders” têm sido alimentada pelo governo da cidade de Chicago. Além disso, foram retiradas colunas de baixa relevância para esse estudo e que dificultavam o processamento dos dados, devido ao extenso tamanho em megabytes do arquivo. Foram retiradas as seguintes colunas da base “Crimes (2001-presente)”: *Case number, Description, Beat, Ward, FBI Code, X Coordinate, Y Coordinate, Updated On, Latitude, Longitude e Location*.

3. Metodologia

As análises foram realizadas em Python, com o auxílio de bibliotecas especializadas para aplicar técnicas de:

- **Estatística descritiva** - Média; mediana; desvio padrão; histogramas; boxplots; distribuição de frequência.
- **Probabilidade** - Cálculos de probabilidade condicional.
- **Inferência estatística** - Testes de hipótese (qui-quadrado e teste t); intervalos de confiança.

Para além disso, as análises deste estudo foram realizadas considerando as *Community Areas* de Chicago, que representam divisões geográficas maiores do que os *Blocks* e, por consequência, são em menor número (77 áreas). Essa abordagem foi adotada para proporcionar uma visualização mais clara e objetiva da distribuição dos crimes e da concentração de criminosos sexuais, evitando a sobrecarga de informações nos gráficos e facilitando a interpretação dos resultados.

Como a base “Sex Offenders” não possuía nenhuma coluna em referência às *Community Areas*, antes de iniciar a análise os criminosos sexuais foram mapeados usando a associação entre os *Blocks* registrados na base de “Sex Offenders” e aqueles presentes na base “Crimes (2001-presente)”, que já contém a identificação das *Community Areas*. Nos casos em que os *Blocks* dos criminosos sexuais não estavam presentes na base de crimes, o mapeamento foi feito manualmente. Para isso foi utilizado o sistema oficial de grade de endereçamento de Chicago, em conjunto com ferramentas como Google Maps e o mapa oficial das *Community Areas* de Chicago.

4. Análises e Resultados

Neste tópico são apresentadas as análises realizadas e resultados obtidos.

4.1. Perfil e Distribuição dos Criminosos Sexuais Registrados

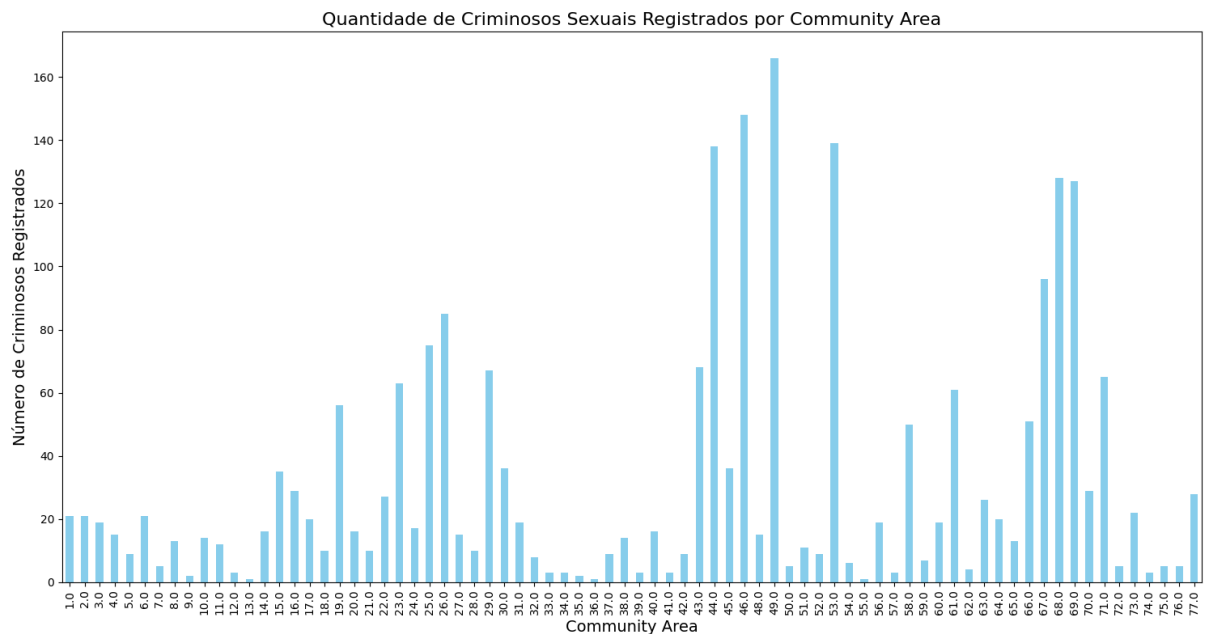
4.1.1. Quantidade de criminosos sexuais registrados por *Community Area*

- **Pergunta:** Qual é a quantidade de criminosos sexuais em cada “Community Area”? Existe um padrão nessa distribuição?
- **Método utilizado:**
 - Gráfico de barras: visualizar a distribuição dos criminosos sexuais pelas ‘Community Areas’.
 - QQ-Plot: visualizar a normalidade da distribuição.
 - Teste de Shapiro Wilk.
- **Descrição:** Calculou-se a frequência absoluta de criminosos sexuais em cada “Community Area”, representando os valores em um gráfico de barras para melhor visualização e entendimento da distribuição deles. Já para entender o comportamento da distribuição desses criminosos pelas “Community Areas” de Chicago foi construído um QQ-plot e aplicado o Teste de Shapiro Wilk.
 - **Hipóteses para o Teste de Shapiro Wilk:**
 - Hipótese Nula (H_0): A distribuição dos criminosos sexuais pelas “Community Areas” de Chicago é normal.

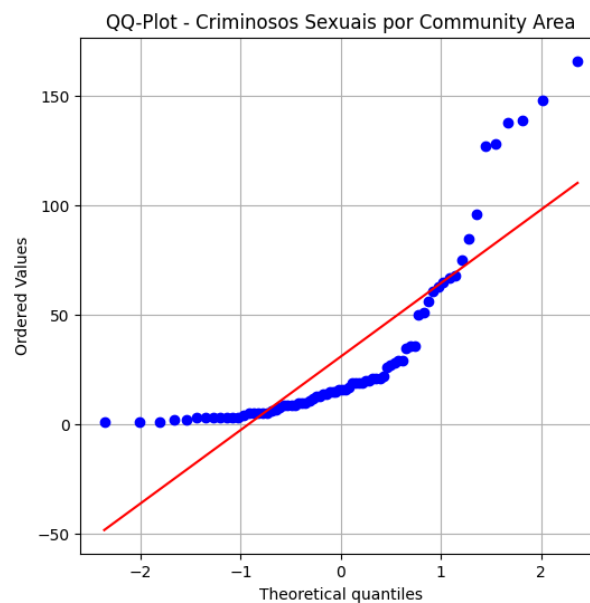
- Hipótese Alternativa (H_1): A distribuição dos criminosos sexuais pelas “Community Areas” de Chicago não é normal.

- **Resultados:**

1. Gráfico de barras



2. QQ-Plot



3. Teste de Normalidade (Shapiro-Wilk):

Estatística: 0.7146, p-valor: 7.365470863511371e-11

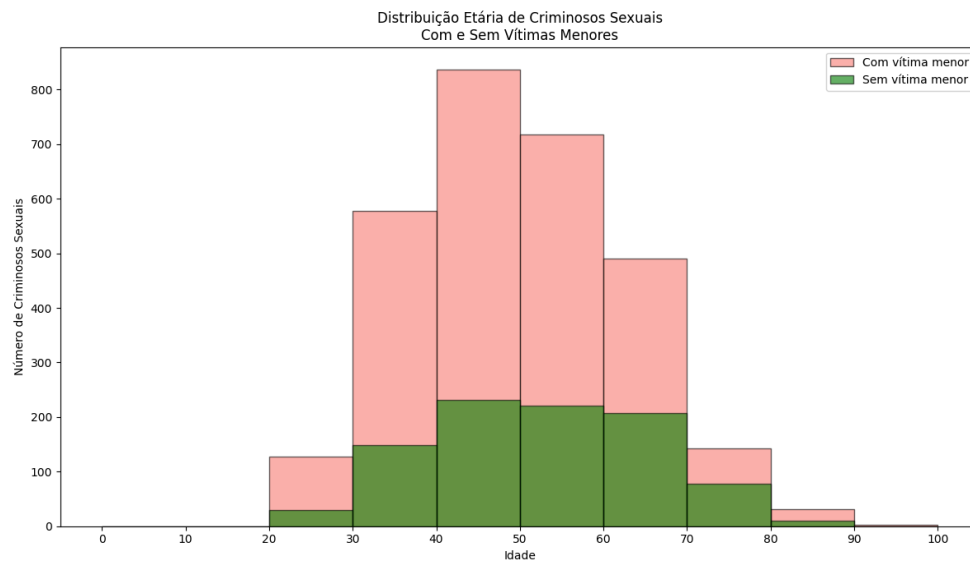
- **Interpretação:** O gráfico de barras e o QQ-plot evidenciam que os infratores sexuais não estão distribuídos de forma equilibrada entre as áreas. A maioria dos bairros registra poucos casos, enquanto um pequeno grupo concentra até mais de 150 infratores. Já o teste de Shapiro-Wilk confirma a informação, considerando que o p-valor é muito baixo, ou seja, a distribuição não é normal. Em resumo, poucos bairros concentram a maior parte dos infratores, sugerindo que características locais influenciam essa distribuição.
-

4.1.2. Distribuição etária dos criminosos sexuais com e sem vítimas menores

- **Pergunta:** Qual a idade dos criminosos sexuais registrados com vítima menor em relação aos registrados sem vítima menor?
- **Método utilizado:**
 - Cálculo da idade dos criminosos sexuais registrados.
 - Frequência absoluta: criminosos sexuais onde a coluna "VICTIM MINOR" é preenchida por 'Y'.
 - Frequência absoluta: criminosos sexuais onde a coluna "VICTIM MINOR" é preenchida por 'N'.
 - Histograma sobreposto: visualizar distribuição etária de criminosos sexuais com e sem vítima menor.
 - Tabela de Contingência: aplicação do Teste Qui-Quadrado.
- **Descrição:** Primeiramente foram calculadas as idades dos criminosos utilizando a data atual (julho-2025) como parâmetro, para conseguir distribuí-los em grupos por faixa etária, facilitando a análise. Posteriormente, foram empregados histogramas sobrepostos para visualizar a distribuição por faixa etária dos criminosos que cometeram crimes contra vítimas menores em comparação daqueles cujas vítimas eram maiores de idade. Para a constatação da existência de alguma associação entre a faixa etária (idade) do criminoso e o fato da vítima ser menor de idade foi aplicado o teste Qui-Quadrado.
 - **Hipóteses para o Teste Qui-Quadrado:**
 - Hipótese Nula (H_0): Não há associação entre a faixa etária do infrator sexual e o fato da vítima ser menor.
 - Hipótese Alternativa (H_1): Há associação entre a faixa etária e o fato da vítima ser menor.

- **Resultados:**

1. Histograma



2. Tabela de Contingência (Faixa Etária x Vítima Menor):

Faixa etária	Victim Minor	
	N	Y
20-29	29	127
30-39	148	578
40-49	231	836
50-59	220	717
60-69	208	491
70-79	78	142
80+	10	34

3. Teste do Qui-Quadrado:

Estatística Qui² = 39.412884890190725

p-valor = 5.940096604151543e-07

Graus de liberdade = 6

- Interpretação:** O histograma e a tabela de contingência indicam que os infratores que atacam menores concentram-se principalmente entre 30 e 59 anos, com pico na faixa de 40–49, enquanto os que têm vítimas adultas

aparecem em menor quantidade e se distribuem de forma mais dispersa. No teste qui-quadrado, o p-valor é baixo (confirmando hipótese alternativa), mostrando que a idade do infrator está associada ao fato de a vítima ser menor. Em resumo, criminosos de meia-idade se relacionam fortemente em crimes contra menores, já contra adultos não há uma relação.

4.1.3. Perfil demográfico (raça, gênero) dos infratores sexuais com e sem vítimas menores

- **Pergunta:** Qual é a raça e o gênero dos infratores sexuais que tiveram vítimas menores? Existem diferenças notáveis nesses perfis em comparação com infratores que não tiveram vítimas menores?
- **Método utilizado:**
 - Frequência absoluta: criminosos sexuais que tiveram vítimas menores divididos por gênero.
 - Frequência absoluta: criminosos sexuais que tiveram vítimas menores divididos por raça.
 - Tabela de contingência: comparar a distribuição percentual dos criminosos sexuais que tiveram vítimas menores por gênero e raça.
 - Frequência absoluta: criminosos sexuais que não tiveram vítimas menores por gênero.
 - Frequência absoluta: criminosos sexuais que não tiveram vítimas menores por raça.
 - Tabela de contingência: comparar a distribuição percentual dos criminosos sexuais que não tiveram vítimas menores por gênero e raça.
 - Tabela de contingência: aplicar teste Qui-Quadrado de independência.
- **Descrição:** Para entender os perfis demográficos dos criminosos, as frequências absolutas deles foram dispostas em dois grupos: com vítima menor e sem vítima menor. A partir disso, foi construída uma tabela de contingência para cada grupo a fim de comparar a distribuição por gênero e raça desses criminosos. Por fim, foi aplicado o teste Qui-Quadrado, para analisar se havia uma diferença significativa entre os perfis desses grupos.
 - **Hipóteses para Teste Qui-Quadrado:**
 - Hipótese Nula (H_0): Não há diferença notável nos perfis demográficos de criminosos com vítima menor e criminosos sem vítima menor.
 - Hipótese Alternativa (H_1): Há diferença notável nos perfis demográficos de criminosos com vítima menor e criminosos sem vítima menor.

- **Resultados:**

1. Frequência Absoluta - Criminosos Sexuais COM Vítima Menor

GENDER:

MALE	2865
FEMALE	60

RACE

BLACK	1516
WHITE	753
WHITE HISPANIC	599
ASIAN/PACIFIC ISLANDER	30
BLACK HISPANIC	16
UNKNOWN / REFUSED	8
AMER INDIAN / ALASKAN NATIVE	3

2. Tabela de contingência (Gênero x Raça) - Criminosos Sexuais COM Vítimas Menores %

	AMER INDIAN / ALASKAN NATIVE	ASIAN / PACIFIC ISLANDER	BLACK	BLACK HISPANIC	UNKNOWN / REFUSED	WHITE	WHITE HISPANIC
FEMALE	0.00	0.00	55.00	0.00	0.00	33.33	11.67
MALE	0.10	1.05	51.76	0.56	0.28	25.58	20.66

3. Frequência Absoluta - Criminosos Sexuais SEM Vítima Menor

GENDER:

MALE	918
FEMALE	6

RACE

BLACK	633
WHITE	161

WHITE HISPANIC	113
ASIAN/PACIFIC ISLANDER	11
BLACK HISPANIC	5
UNKNOWN / REFUSED	1
AMER INDIAN / ALASKAN NATIVE	0

4. Tabela de contingência (Gênero x Raça) - Criminosos Sexuais SEM Vítimas Menores %

	AMER INDIAN / ALASKAN NATIVE	ASIAN / PACIFIC ISLANDER	BLACK	BLACK HISPANIC	UNKNOWN / REFUSED	WHITE	WHITE HISPANIC
FEMALE	0.00	0.00	33.33	0.00	0.00	33.33	33.33
MALE	0.00	1.20	68.74	0.54	0.11	17.32	12.09

5. Teste Qui-Quadrado (Gênero x Vítima Menor):

Estatística Qui²: 7.3781, p-valor: 0.0066, graus de liberdade: 1

6. Teste Qui-Quadrado (Raça x Vítima Menor):

Estatística Qui²: 83.2244, p-valor: 0.0000, graus de liberdade: 6

- **Interpretação:** As contagens mostram que, embora a maioria dos infratores seja do sexo masculino nos dois grupos, a proporção de mulheres com vítimas menores é levemente maior (cerca de 2%) do que entre os que agrediram apenas adultos (menos de 1%). Já no recorte racial, infratores com vítimas menores são em maioria brancos e brancos hispânicos, enquanto os que agrediram adultos são de maioria negros. O teste qui-quadrado evidencia que o p-valor de gênero é 0,0066 e sobre a raça, p-valor < 0,001, rejeitando a hipótese de perfis demográficos iguais nos dois grupos. Em síntese, existem diferenças estatisticamente significativas tanto de gênero quanto de raça entre infratores com vítimas menores e adultos, indicando que os perfis demográficos variam conforme o tipo de vítima.

4.1.4. Faixas etárias mais associadas a vítimas menores e distribuição geográfica

- **Pergunta:** Quais faixas etárias de criminosos sexuais estão mais associadas a vítimas menores e como elas se distribuem por “Community Area”?

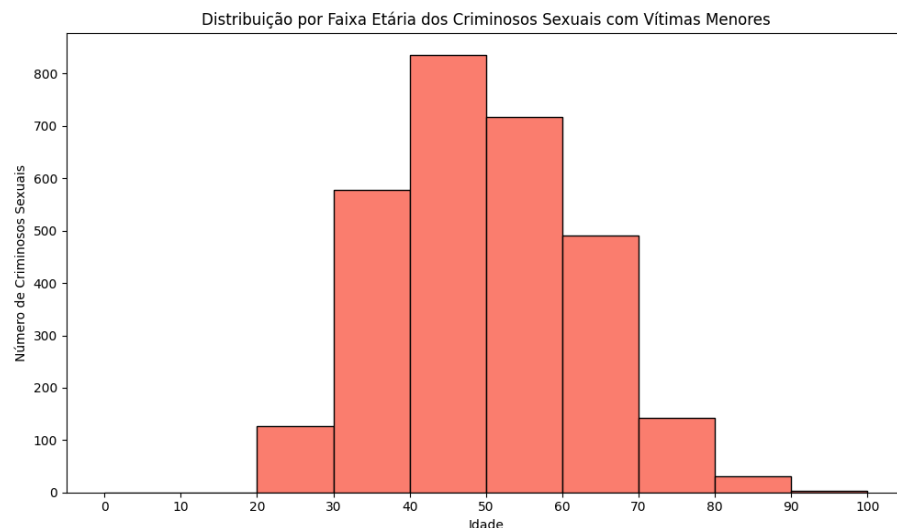
- **Método utilizado:**

- Cálculo da idade dos criminosos sexuais registrados.
- Separação de criminosos sexuais onde a coluna 'VICTIM MINOR' é preenchida por “y”.
- Histograma: distribuição dos criminosos sexuais com vítima menor de idade , dentro das faixas de 10 a 10 anos.
- Gráfico de barras: quantidade de criminosos por “Community Area”.

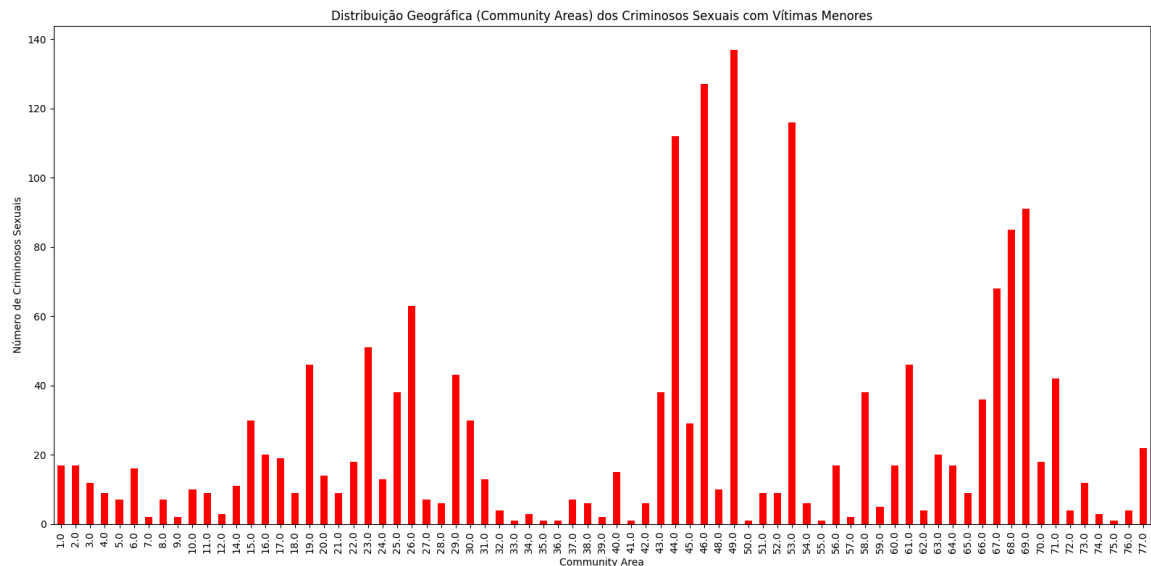
- **Descrição:** A fim de identificar as faixas etárias predominantes entre dos crimes contra vítimas menores de idade, foi calculada a idade dos criminosos sexuais a partir das suas datas de nascimento, presentes na coluna “BIRTH DATE”. Então, para melhor visualização e análise do resultado, foi construído um histograma. Por fim, foi construído um gráfico de barras para análise da distribuição geográfica (“Community Area”) desses criminosos.

- **Resultados:**

1. Histograma



2. Gráfico de barras



- **Interpretação:** O histograma revela que a maior parte dos infratores com vítimas menores se concentra entre 40 e 49 anos, seguida pelas faixas de 50–59 e 30–39. Já o gráfico de barras indica que esses mesmos infratores não estão espalhados por toda a cidade, eles se agrupam em “Community Areas” de regiões do centro-sul e do extremo sul. Em resumo, crimes contra menores estão associados principalmente a infratores de meia-idade e tendem a ocorrer repetidamente nos mesmos bairros, sugerindo que fatores demográficos e locais contribuem simultaneamente para essa concentração.

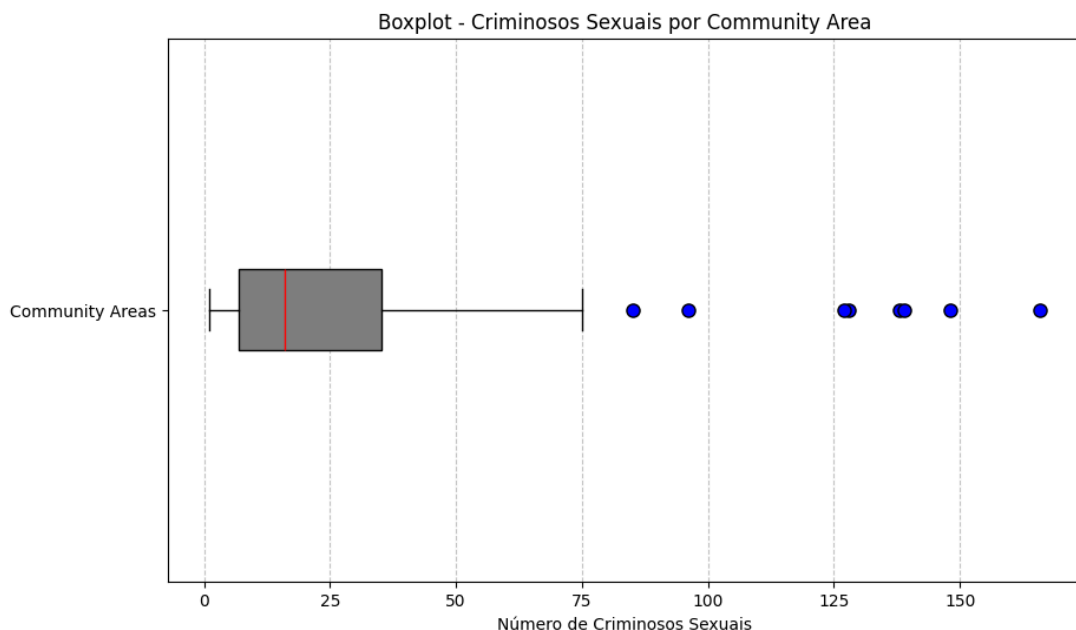
4.1.5. Crimes prevalentes e perfis demográficos nas áreas com mais criminosos sexuais

- **Pergunta:** Quais são as "Community Areas" com maior número de criminosos sexuais? Quais tipos de crimes são mais comuns nessas áreas? Qual o perfil demográfico (raça, gênero) dos infratores que vivem nessas áreas?
- **Método utilizado:**
 - Frequência absoluta: criminosos sexuais por "Community Area" a partir dos dados válidos (remoção de valores nulos).
 - Média, moda e mediana: número de criminosos por “Community Area”.
 - Coeficiente de Variação (CV): criminosos por área.
 - Boxplot: contagem de criminosos sexuais por “Community Area”.
 - Gráfico de barras: visualizar a distribuição de criminosos sexuais em "Community Areas" com alta concentração de criminosos sexuais registrados.

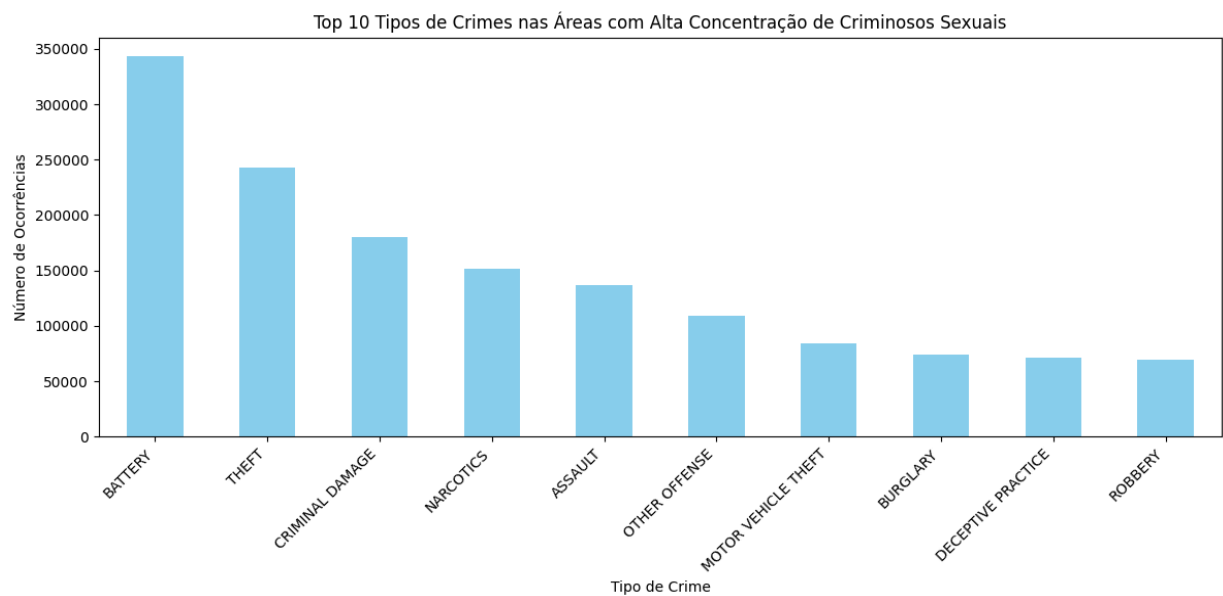
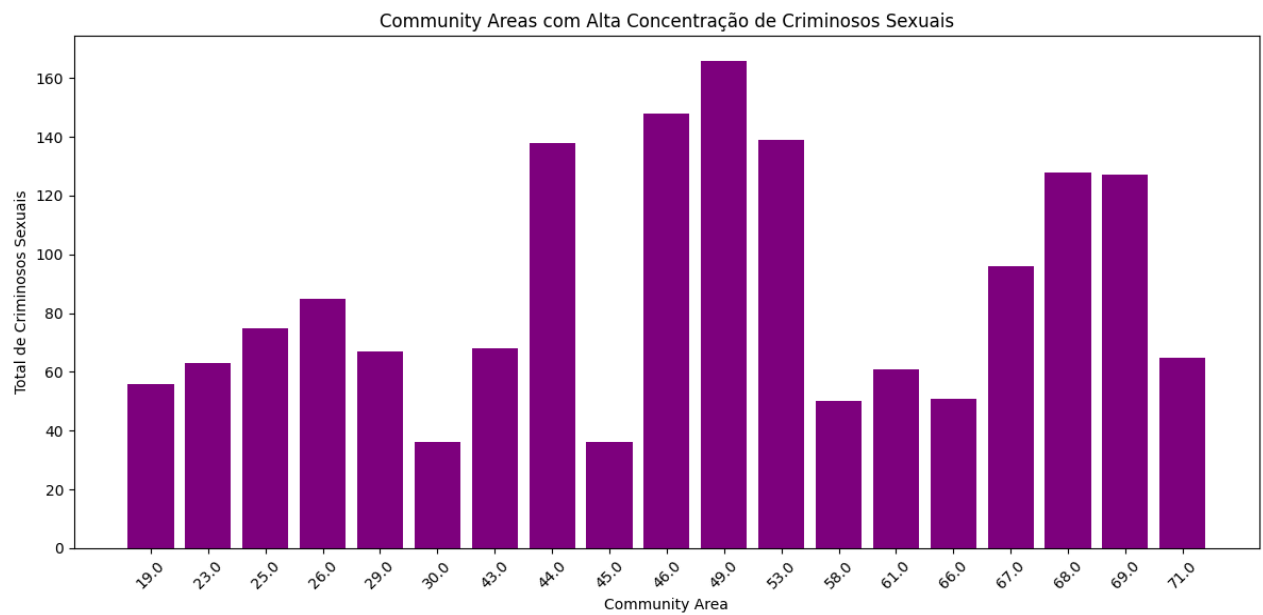
- Gráfico de barras: 10 tipos de crimes mais frequentes nas áreas com alta concentração de criminosos sexuais.
- Tabela de contingência: cruzamento gênero e raça dos criminosos sexuais nas “Community Areas” e suas distribuições percentuais.
- **Descrição:** Primeiro, calculou-se média, mediana e moda de criminosos sexuais por área. Então, foi construído um boxplot para identificar possíveis desbalanceamentos nessa distribuição. Logo após, o CV foi calculado para entender a variação relativa do número de criminosos sexuais entre as “Community Areas” e, a partir disso, determinou-se a quantidade de criminosos considerada “maior/alta” (foram consideradas áreas com maior número de criminosos sexuais aquelas que o número de criminosos fosse maior que o terceiro quartil). Então, construiu-se um gráfico de barras para melhor visualização da relação entre áreas com alta concentração de infratores sexuais e tipos de crimes mais cometidos nelas. Por fim, a tabela de contingência foi utilizada para entender a distribuição dos criminosos por gênero e raça.
- **Resultados:**
 1. Média de criminosos por Community Area: 31.06578947368421
Moda de criminosos por Community Area: 3
Mediana de criminosos por Community Area: 16.0

2. Coeficiente de Variação (CV): 125.38%

3. Boxplot



4. Gráficos de barras



5. Frequência absoluta

GENDER

MALE	1634
FEMALE	21

RACE

BLACK	963
WHITE	377
WHITE HISPANIC	294
BLACK HISPANIC	10

ASIAN / PACIFIC ISLANDER	7
UNKNOWN / REFUSED	3
AMER INDIAN / ALASKAN NATIVE	1

6. Tabela de contingência

	AMER INDIAN / ALASKAN NATIVE	ASIAN / PACIFIC ISLANDER	BLACK	BLACK HISPANIC	UNKNOWN / REFUSED	WHITE	WHITE HISPANIC
FEMALE	0.00	0.00	57.14	0.00	33.33	9.52	0.00
MALE	0.06	0.43	58.20	0.61	22.64	17.87	0.18

- **Interpretação:** O boxplot, aliado aos valores de média (31), mediana (16) e ao coeficiente de variação de 125 %, mostram que a maioria das “Community Areas” tem poucos infratores sexuais, sendo que apenas um grupo pequeno de bairros reúne entre 70 e 130 indivíduos, demonstrando forte desigualdade na distribuição. Nessas áreas de alta concentração, os dez crimes mais frequentes são patrimoniais e de violência geral, sugerindo que a presença de infratores sexuais faz parte de um contexto mais amplo de criminalidade elevada. O perfil demográfico desses infratores é predominantemente masculino e de raça predominantemente negra ($\approx 59\%$), branca ($\approx 23\%$) e branca hispânica ($\approx 18\%$).

4.2. Análise Geográfica e Comparação Entre Áreas

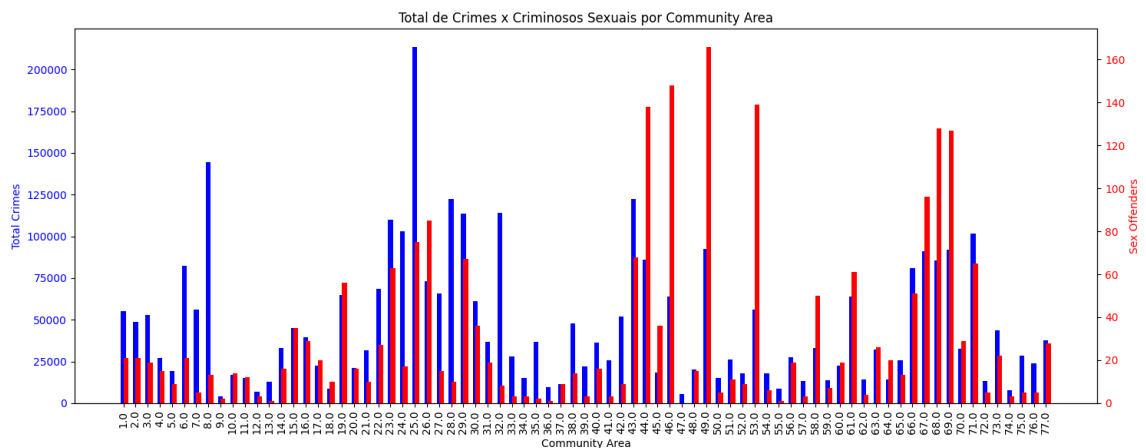
4.2.1. Community Areas com alta frequência de crimes gerais e alta concentração de criminosos sexuais

- **Pergunta:** Existem "Community Areas" em Chicago com alta frequência de crimes gerais e alta concentração de criminosos sexuais simultaneamente?
- **Método utilizado:**
 - Frequência absoluta: crimes gerais por “Community Area”.
 - Frequência absoluta: criminosos sexuais por “Community Area”.
 - Gráfico de barras comparativo: crimes gerais e criminosos sexuais.
- **Descrição:** Antes de tudo, foram removidos da base crimes e criminosos sem “Community Area” registrada. Para identificar as “Community Areas” com altas incidências, foi feita a contagem das frequências absolutas tanto dos crimes quanto dos criminosos sexuais em cada “Community Area”. Então, foi construído um gráfico de barras comparativo para melhor

visualização e análise das diferenças de frequência e possíveis relações entre esses dois indicadores em cada “Community Area”.

- **Resultados:**

1. Gráfico de barras comparativo:



- **Interpretação:** O gráfico evidencia que crimes gerais existem em quase todas as “Community Areas”, mas apenas um subconjunto delas apresenta picos simultâneos nas duas barras: as áreas 24, 25, 28 e 44–49. Em resumo, existem sim áreas que reúnem alta criminalidade geral e grande quantidade de infratores sexuais ao mesmo tempo, o que sugere maior foco de políticas de segurança nessas localidades.

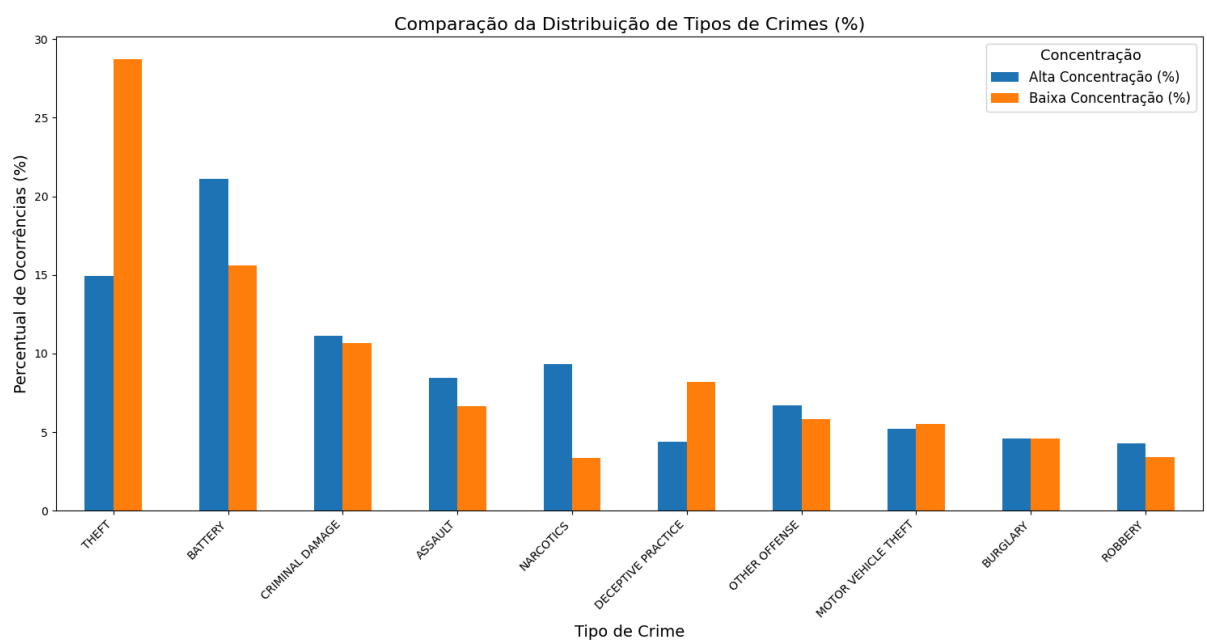
4.2.2. Diferença na distribuição dos tipos de crimes entre áreas com alta e baixa concentração de criminosos sexuais

- **Pergunta:** Os principais tipos de crimes são diferentes entre as “Community Areas” com alta e com baixa concentração de criminosos sexuais?
- **Método utilizado:**
 - Frequência absoluta: crimes nas áreas de alta concentração de criminosos sexuais registrados.
 - Frequência absoluta: crimes nas áreas de baixa concentração de criminosos sexuais registrados.
 - Frequência relativa: crimes por “Primary Type” nas áreas de alta concentração de criminosos sexuais registrados.
 - Frequência relativa: crimes por "Primary Type" nas áreas de baixa concentração de criminosos sexuais registrados.
 - Gráfico de barras comparativo: visualização das distribuições dos crimes por "Primary Type" nas áreas de alta e baixa concentração de criminosos sexuais registrados.

- **Descrição:** Primeiramente, as “Community Areas” foram separadas em dois grupos: com maior ou menor concentração de criminosos sexuais (considerações: áreas de alta concentração -> frequência absoluta de criminosos maior que o terceiro quartil; áreas de baixa concentração -> frequência absoluta de criminosos menor que o terceiro quartil). Para comparação entre as frequências dos tipos criminais, foi construído um gráfico de barras com os dez tipos criminais mais recorrentes nas duas categorias.

- **Resultados:**

1. Gráfico de barras



2. Frequência absoluta- top 10 maiores e menores concentrações sex offenders

Primary Type	Community Area	
	Alta concentração	Baixa concentração
ASSAULT	136826	133310
BATTERY	343023	312367
BURGLARY	74181	92371
CRIMINAL DAMAGE	180296	213063
DECEPTIVE PRACTICE	71194	163625

MOTOR VEHICLE THEFT	84172	110925
NARCOTICS	151380	66802
OTHER OFFENSE	108817	116121
ROBBERY	69391	68213
THEFT	242528	574297

- **Interpretação:** O gráfico comparativo mostra que nas Community Areas que concentram mais infratores sexuais, os crimes principais são violência (assault) e tráfico (battery). Já nas áreas de baixa concentração, dominam delitos patrimoniais (theft, deceptive practice, burglary e motor-vehicle theft). Em ambas as categorias battery permanece no topo da lista absoluta, mas cresce onde há mais infratores sexuais. Em síntese, os infratores sexuais tendem a residir em regiões onde crimes violentos e ligados a drogas são mais comuns, ao passo que bairros com poucos desses infratores possuem mais crimes de patrimônio.

4.2.3. Comparação dos três principais crimes em áreas de alta e baixa incidência de crimes sexuais

- **Pergunta:** Quais são os três tipos de crimes gerais mais comuns nas “Community Areas” com maior incidência de crimes sexuais, e como eles se comparam com as áreas de baixa incidência de crimes sexuais? Há diferença estatisticamente significativa na distribuição desses crimes entre as duas categorias de áreas?
- **Método utilizado:**
 - Frequência absoluta: crimes sexuais por “Community Areas”.
 - Seleção: crimes apenas das áreas extremas (baixa ou alta incidência).
 - Frequência relativa: tipos de crimes gerais nas duas área extremas.
 - Seleção dos 3 tipos de crimes mais comuns nas áreas de alta incidência.
 - Seleção dos 3 tipos de crimes mais comuns nas áreas de baixa incidência.
 - Gráfico de barras: tipo de crime por número de ocorrência.
 - Tabela de contingência: crime por área.
 - Teste Qui-Quadrado.
- **Descrição:** Primeiro, as “Community Areas” foram separadas em dois grandes grupos: com maior ou menor incidência de crimes sexuais

(considerações: áreas de alta incidência -> as cinco “Community Areas” com maior frequência absoluta de registros de crimes sexuais; áreas de baixa incidência -> as cinco “Community Areas” com menor frequência absoluta de registros de crimes sexuais). A partir disso, foi identificado os três tipos de crimes que mais ocorreram em cada categoria. Dado que os três tipos de crimes que mais ocorreram foram os mesmos para as duas categorias, construiu-se uma de uma tabela de contingência e aplicado o teste Qui-Quadrado para identificar se a frequência relativa é estaticamente parecida nas duas categorias.

- **Hipóteses para o Teste Qui-Quadrado:**

- H_0 (hipótese nula): Não há associação entre o tipo de crime e a categoria da área (alta ou baixa incidência de crimes sexuais).
- H_1 (hipótese alternativa): Existe associação entre o tipo de crime e a categoria da área.

- **Resultados:**

1. Áreas com alta incidência: [25, 8, 28, 43, 29]
Áreas com baixa incidência: [37, 18, 12, 47, 9]

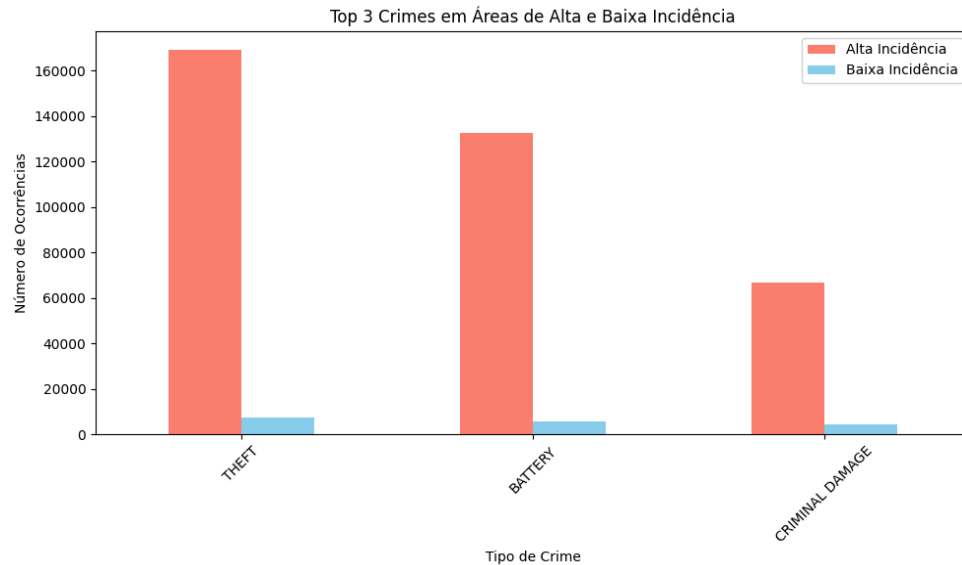
2. Top 3 crimes em áreas de alta incidência:

THEFT	168879
BATTERY	132562
CRIMINAL DAMAGE	66988

3. Top 3 crimes em áreas de baixa incidência:

THEFT	7408
BATTERY	5872
CRIMINAL DAMAGE	66988

4. Gráfico de barras comparativo



5. Tabela de Contingência:

Primary type	THEFT	BATTERY	CRIMINAL DAMAGE
Alta incidência	168879	132562	66988
Baixa incidência	7408	5872	4357

6. Resultado do Teste Qui-Quadrado:

Estatística Qui-Quadrado: 475.470570687796

p-valor: 5.660794919431076e-104

Graus de Liberdade: 2

- **Interpretação:** O gráfico de barras evidencia que, tanto nas cinco áreas com mais crimes sexuais quanto nas cinco com menos, os delitos mais comuns são os mesmos (theft, battery e criminal damage). No entanto, o volume desses crimes é muito maior nas áreas de alta incidência. A tabela de contingência, testada com qui-quadrado, confirma que essa diferença tem relação significativa dado o p-valor pequeno, rejeitando a hipótese nula. Em resumo, embora o ranking dos três principais crimes permaneça igual, sua frequência relativa muda de forma estatisticamente significativa, indicando que áreas com muitos crimes sexuais também registram proporções mais altas de delitos.

4.3. Análises Temporais

4.3.1. Relação entre a tendência dos crimes sexuais e crimes gerais ao longo dos anos

- **Pergunta:** Há relação entre a tendência do número de crimes sexuais com o número de crimes gerais ocorridos ao longo dos anos?
- **Método utilizado:**
 - Seleção de crimes por categoria: sexuais e não sexuais.
 - Frequência absoluta: ocorrências por ano e categoria.
 - Coeficiente de correlação linear de Pearson
 - Gráfico de linhas comparativo: visualizar a distribuição das ocorrências de crimes sexuais e crimes não sexuais por ano.
- **Descrição:** A partir da coluna “Primary Type”, os crimes foram separados em dois tipos: sexuais e não sexuais, a fim de calcular as frequências absolutas das ocorrências de cada tipo por ano. Então, para entender se havia alguma correlação de força e direção, entre as tendências de ocorrências ao longo dos anos por esses tipos criminais, foi aplicado o Coeficiente de Correlação de Pearson (r). Por fim, foi construído um gráfico de linhas comparativo para melhor visualização dos dados, uma vez que a quantidade de crimes não sexuais está na casa das centenas de milhares, mas a de crimes sexuais está na casa de milhares, podendo afetar a comparação se fosse um gráfico de linhas comum (com apenas um eixo y).
- **Resultados:**

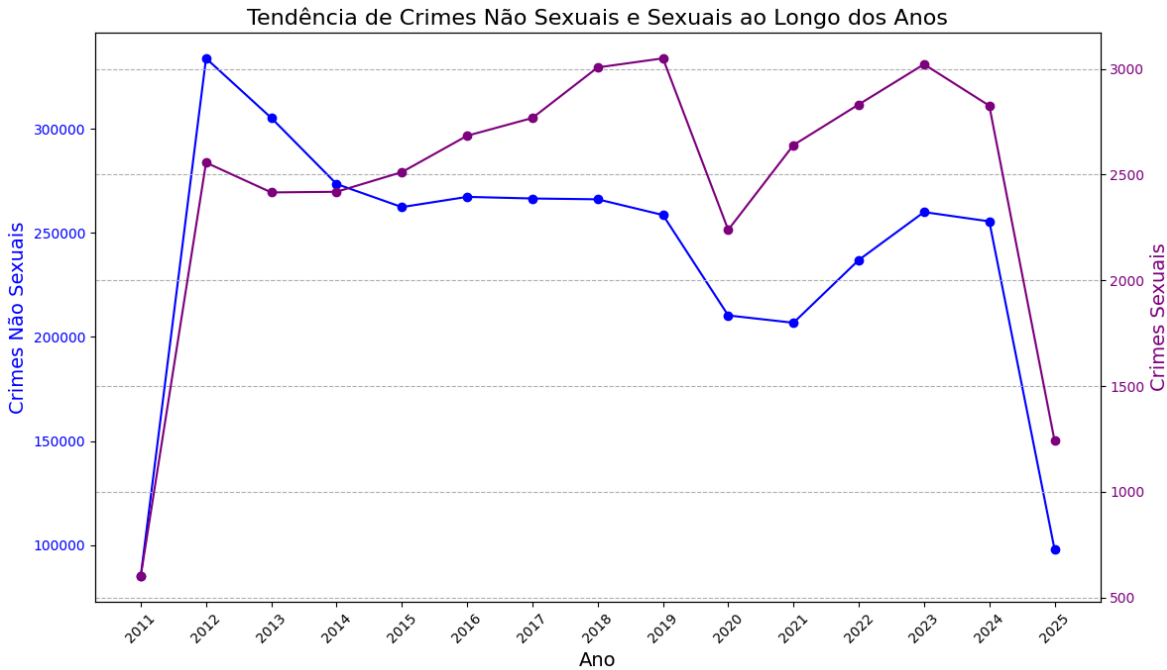
1. Frequência absoluta

ano	crimes não sexuais	crimes sexuais
2011	85338	603
2012	333769	2557
2013	305160	2416
2014	273443	2419
2015	262359	2512
2016	267263	2684
2017	266490	2768

2018	266091	3007
2019	258564	3050
2020	210349	2239
2021	206828	2640
2022	206828	2832
2023	236963	3022
2024	259972	2826
2025	98023	1242

2. Correlação entre crimes não sexuais e crimes sexuais ao longo dos anos (r):
 $r = 0.82$

3. Gráfico de linhas comparativo



- **Interpretação:** O gráfico de linhas mostra que os crimes sexuais e não sexuais se movem em sincronia. O coeficiente de correlação de Pearson reforça essa percepção, indicando associação positiva forte. Em resumo, quando o total de crimes gerais aumenta ou diminui, os crimes sexuais tendem a acompanhar a mesma direção, sugerindo que fatores externos que afetam a criminalidade como um todo também influenciam as ocorrências sexuais.

4.3.2. Variação anual da proporção de crimes sexuais sobre o total de crimes

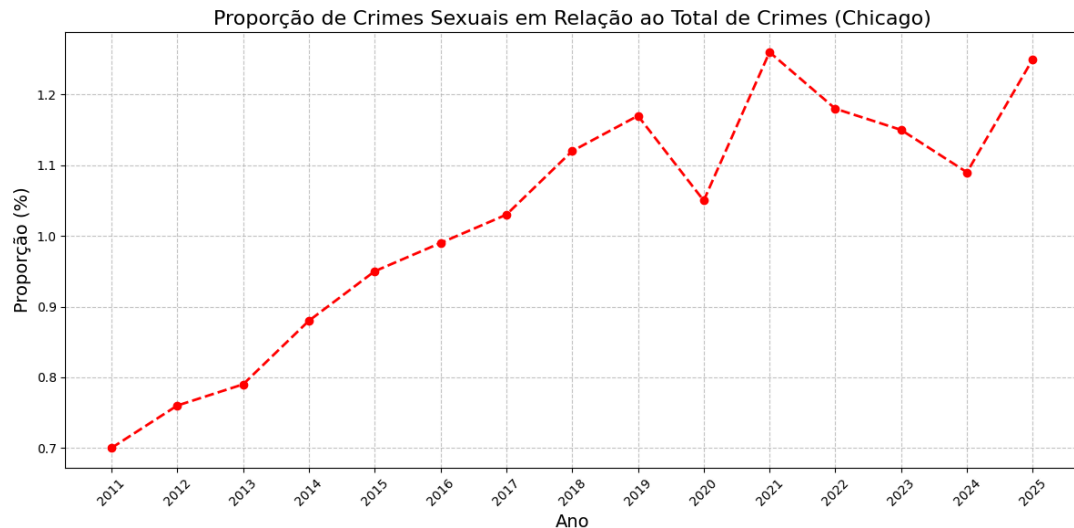
- **Pergunta:** Qual a variação dos crimes sexuais em relação ao total de crimes anualmente em Chicago?
- **Método utilizado:**
 - Frequência absoluta: crimes sexuais e crimes gerais por ano.
 - Frequência relativa: total de crimes sexuais pelo total de crimes gerais, por ano.
 - Gráfico de linha: visualizar as proporções ao longo dos anos.
- **Descrição:** Foram selecionados apenas os crimes de “Primary type” caracterizados como sexuais, e calculada a frequência absoluta deles para o cálculo de proporção em relação ao total de crimes por ano. Então, foi construído um gráfico de linha para facilitar a visualização e interpretação dessas proporções ao longo dos anos presentes (2011 a 2025).
- **Resultados:**

1. Tabela de contingência

ano	crimes gerais	crimes sexuais	proporção (%)
2011	85941	603	0.70
2012	336326	2557	0.76
2013	307576	2416	0.79
2014	275862	2419	0.88
2015	264871	2512	0.95
2016	269947	2684	0.99
2017	269258	2768	1.03
2018	269098	3007	1.12
2019	261614	3050	1.17
2020	212588	2239	1.05
2021	209468	2640	1.26
2022	239795	2832	1.18
2023	262994	3022	1.15
2024	258332	2826	1.09

2025	99265	1242	1.25
------	-------	------	------

2. Gráfico de linha



- **Interpretação:** O gráfico de linha mostra que a participação dos crimes sexuais no total de ocorrências sobe quase de forma contínua, diminuindo consideravelmente apenas em 2020 e 2024. Então, é possível sugerir que o avanço das proporções é consistente (a cada 5 anos os delitos sexuais crescem aproximadamente meio percentual), evidenciando que eles aumentam mais rápido que a criminalidade geral. Em síntese, os crimes sexuais vêm ganhando peso relativo no panorama criminal de Chicago.

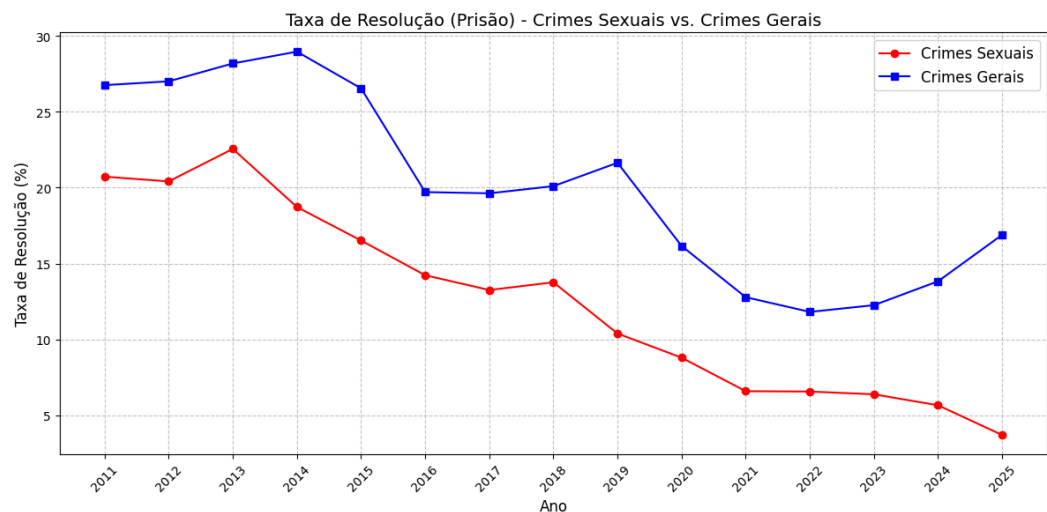
4.3.3. Taxa de resolução (prisão) de crimes sexuais versus crimes gerais ao longo dos anos

- **Pergunta:** Existe alguma relação entre a taxa de prisão para crimes sexuais e para crimes gerais?
- **Método utilizado:**
 - Seleção dos registros por tipo de crime (sexual ou não sexual).
 - Frequência absoluta: crimes reportados e resolvidos de cada tipo por ano.
 - Frequência relativa: taxa de resolução por tipo de crime (sexual ou não sexual).
 - Tratamento de casos com zero ocorrências para evitar divisão por zero.
 - Gráficos de linhas: visualizar proporções anuais de cada tipo de crime.
 - QQ-plot: avaliar possível normalidade das distribuições.

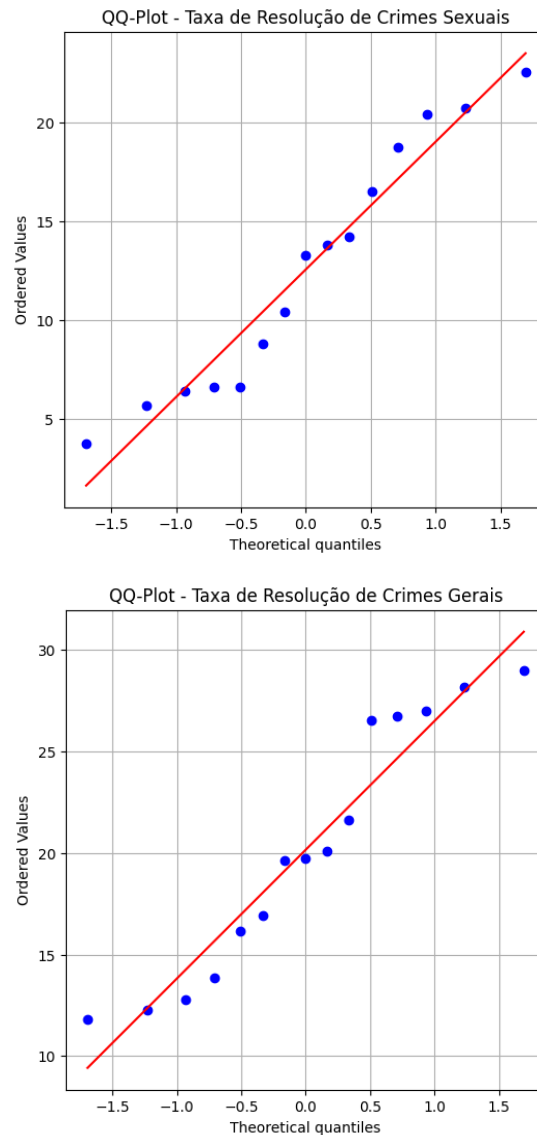
- Teste de Shapiro Wilk.
- Teste T.
- **Descrição:** Primeiramente, os crimes foram separados entre sexuais e não sexuais. Depois, foi feita a frequência absoluta de crimes reportados e de crimes resolvidos a fim de calcular a taxa de resolução de cada categoria por ano e, então, conseguir estabelecer uma comparação (evitando divisões nulas). Logo após, um gráfico de linhas comparativo foi construído para melhor análise visual. A fim de analisar uma possível diferença realmente significativa entre as taxas, foram construídos dois QQ-plots. Por fim, para confirmar a existência de uma distribuição normal, foi aplicado o teste de Shapiro Wilk. Após a confirmação, foi aplicado o Teste T para analisar a significância.
 - **Hipóteses para o Teste T:**
 - H_0 (nula): Não há diferença significativa nas taxas médias de resolução entre crimes sexuais e crimes gerais.
 - H_1 (alternativa): Há diferença significativa nas taxas médias de resolução entre crimes sexuais e crimes gerais.

- **Resultados:**

- 1. Gráfico de linhas



2. QQ-Plots



3. Análise de Normalidade da Taxa de Resolução de Crimes Sexuais:
ShapiroResult(statistic=np.float64(0.9304573386292063),
pvalue=np.float64(0.2772026362229164))

4. Análise de Normalidade da Taxa de Resolução de Crimes Gerais:
ShapiroResult(statistic=np.float64(0.9081135358327677),
pvalue=np.float64(0.12668456361821367))

5. Teste t pareado:
Estatística t: -12.067836533408418
p-valor: 8.700141447429958e-09

- **Interpretação:** O gráfico de linhas evidencia que as duas curvas descem ao longo do tempo, mas mantendo percentuais distintos. Já os QQ-plots indicam que as duas variáveis seguem uma distribuição normal (considerando que os

p-valores são maiores que 0,05). Então, o teste t pareado confirma a hipótese alternativa, onde há diferença média significativa entre as duas taxas, embora os números de crimes gerais sejam bem mais altos que os sexuais. Em suma, embora as duas variáveis tenham tendência de queda, é sugerido que resolver crimes sexuais é mais difícil do que resolver delitos em geral.

4.4. Testes de Hipótese e Inferência Estatística

4.4.1. Diferença na média anual de crimes entre áreas com alta e baixa concentração de criminosos sexuais.

- **Pergunta:** A média anual de crimes é diferente entre as “Community Areas” com uma proporção maior de criminosos sexuais e aqueles com uma proporção menor?
- **Método utilizado:**
 - Frequência absoluta: áreas de baixa concentração de criminosos sexuais (menor ou igual que os registrados no terceiro quartil).
 - Média: crimes registrados nas áreas de baixa concentração por ano.
 - Frequência absoluta: áreas de alta concentração de criminosos sexuais (maior que os registrados no terceiro quartil).
 - Média: crimes registrados nas áreas de alta concentração por ano.
 - Teste de normalidade (Shapiro).
 - Teste Mann-Whitney U.
- **Descrição:** As “Community Areas” foram divididas entre alta e baixa concentração de criminosos. Após isso, foi calculada a média anual de crimes em cada grupo e aplicado o Teste de Shapiro-Wilk para verificar a normalidade das distribuições. Como umas das categorias apresentaram distribuição normal (“Community Areas” com alta concentração) e a outra não (“Community Areas” com baixa concentração), utilizou-se o Teste de Mann-Whitney U para comparar as médias e entender se havia diferença significativa entre as áreas de alta e baixa concentração de criminosos sexuais registrados, em relação à média anual de crimes.
 - **Hipóteses para o Teste de Mann-Whitney U:**
 - Hipótese Nula (H_0): A distribuição da média anual de crimes nas “Community Areas” com alta concentração de criminosos sexuais é igual à das “Community Areas” com baixa concentração.
 - Hipótese Alternativa (H_1): A distribuição da média anual de crimes nas “Community Areas” com alta concentração é diferente da das “Community Areas” com baixa concentração.

- **Resultados:**

1. Média anual de crimes gerais nas Community Areas com baixa concentração de criminosos sexuais: 132986.93
2. Média anual de crimes gerais nas Community Areas com alta concentração de criminosos sexuais: 108193.47
3. Normalidade - Áreas com baixa concentração:
ShapiroResult(statistic=np.float64(0.7959629831829911),
pvalue=np.float64(0.0032644538889512018))
4. Normalidade - Áreas com alta concentração:
ShapiroResult(statistic=np.float64(0.891362919797989),
pvalue=np.float64(0.07032169455696578))
5. Teste de Mann-Whitney U:
Estatística U = 170.0, p-valor = 0.01806662299891238

- **Interpretação:** O teste de Mann-Whitney U mostra que as distribuições de crimes anuais são diferentes entre áreas de alta e baixa concentração de criminosos sexuais. As "Community Areas" com maior proporção de criminosos sexuais registram, em média, 108 mil delitos por ano, enquanto aquelas com menor proporção alcançam 133 mil (diferença estatisticamente significativa). Em Suma, a existência de muitos infratores sexuais não implica maior volume de criminalidade geral. Na realidade, os bairros com menos desses infratores são justamente os que concentram mais ocorrências totais ao longo do ano.

4.4.2. Probabilidade de ocorrência de crimes sexuais em áreas com mais de um criminoso registrado

- **Pergunta:** A probabilidade de um crime sexual ocorrer em uma "Community Area" na qual existe mais de um criminoso sexual registrado é maior que nas "Community Areas" com um ou nenhum criminoso sexual registrado?
- **Método utilizado:**
 - Frequência absoluta: crimes ocorridos em "Community Areas" que possuem mais de um criminoso sexual registrado.
 - Frequência absoluta: crimes ocorridos em "Community Areas" que possuem um ou menos criminosos sexuais registrados.

- Probabilidade de ocorrer um crime em “Community Areas” com mais de um criminoso sexual registrado.
 - Probabilidade de ocorrer um crime em “Community Areas” com um ou menos criminosos sexuais registrados.
 - Teste Z.
- **Descrição:** A fim de analisar se existe diferença entre as probabilidades de ocorrência, calculou-se a frequência dos crimes nas “Community Areas” pelas diferentes categorias: com mais de um criminoso sexual registrado ou com menos de dois criminosos sexuais registrados.
 - **Hipóteses para o teste Z:**
 - H_0 (nula): As proporções são iguais ou a proporção na área com mais criminosos não é maior.
 - H_1 (alternativa): A proporção em áreas com mais de 1 criminoso é maior que nas demais áreas.
- **Resultados:**
 1. Probabilidade de um crime sexual ocorrer em “Community Areas” com mais de 1 criminoso registrado: 0.99
 2. Probabilidade de um crime sexual ocorrer em “Community Areas” com 1 ou nenhum criminoso registrado: 0.01
 3. Teste Z:
 4. Estatística Z: 266.14520209388434
 5. P-valor: 0.0
 - **Interpretação:** Os números mostram que as “Community Areas” onde residem dois ou mais sex-offenders concentram, em média, uma taxa de crimes sexuais 99 vezes maior do que aquelas que abrigam apenas um ou nenhum infrator registrado, sugerindo forte associação positiva entre a presença de múltiplos criminosos sexuais e a incidência de novos delitos da mesma natureza.

4.4.3. Probabilidade de um crime ocorrer em área com pelo menos um criminoso sexual registrado

- **Pergunta:** Qual a probabilidade de um crime ocorrer em uma “Community Area” que possui um número de criminosos sexuais maior que a mediana?
- **Método utilizado:**

- Frequência absoluta: criminosos sexuais por "Community Area".
 - Mediana: criminosos por "Community Area".
 - Seleção: "Community Areas" com número de criminosos sexuais registrados maior que a mediana.
 - Frequência absoluta: crimes nas "Community Areas" selecionadas.
 - Probabilidade: quantidade de crimes ocorridos em "Community Areas" com mais criminosos sexuais registrados que a mediana pela quantidade total de crimes ocorridos.
- **Descrição:** Em primeiro lugar, foi realizada a contagem do número de criminosos sexuais registrados em cada "Community Area". Calculou-se a mediana do número de criminosos sexuais por "Community Area". Em seguida, foram selecionadas as áreas que apresentavam mais criminosos sexuais registrados que a mediana. Depois foi calculado o total de crimes ocorridos nas "Community Areas" selecionadas, para então, calcular a probabilidade com base na proporção entre o número de crimes registrados nas "Community Areas" com presença de criminosos sexuais e o total de crimes registrados na cidade, representando assim a probabilidade de um crime ocorrer em uma área que possui mais criminosos sexuais registrados que a mediana.
- **Resultados:**
 1. Mediana de criminosos por Community Area: 16.0
 2. Community Areas com mais criminosos sexuais registrados que a mediana: [1.0, 2.0, 3.0, 6.0, 15.0, 16.0, 17.0, 19.0, 22.0, 23.0, 24.0, 25.0, 26.0, 29.0, 30.0, 31.0, 43.0, 44.0, 45.0, 46.0, 49.0, 53.0, 56.0, 58.0, 60.0, 61.0, 63.0, 64.0, 66.0, 67.0, 68.0, 69.0, 70.0, 71.0, 73.0, 77.0]
 3. Número de crimes nessas áreas: 2387955
 4. Probabilidade de ocorrer um crime aleatório em uma 'Community Area' com mais criminosos sexuais registrados que a mediana: 0.6591
 - **Interpretação:** Os números mostram que 65,91% de todos os delitos registrados na cidade ocorreram nas "Community Areas" cujo número de infratores sexuais excede a mediana (16 indivíduos por área). Ou seja, a chance de um crime aleatório ter acontecido em um desses bairros "acima da mediana" é de cerca de dois terços. Isso evidencia que grande parte dos crimes gerais se concentram justamente onde há mais infratores sexuais. No entanto, esta probabilidade é derivada da distribuição dos crimes já ocorridos, então pode refletir fatores como densidade habitacional ou características

socioeconômicas, além de uma relação de causa e consequência entre a presença de infratores sexuais e o volume total de crimes.

4.4.4. Teste sobre a probabilidade de crimes sexuais ser maior em áreas com alta concentração de criminosos sexuais

- **Pergunta:** A probabilidade de um crime sexual ocorrer em uma "Community Area" com alta concentração de criminosos sexuais registrados é significativamente maior do que em uma "Community Area" com baixa concentração de criminosos sexuais registrados?
- **Método utilizado:**
 - Seleção: "Community Areas" com alta (acima do terceiro quartil) e baixa (abaixo do terceiro quartil) concentração de criminosos sexuais.
 - Proporção de crimes sexuais por total de crimes para cada grupo.
 - Teste 'Z': analisar as duas proporções (teste unilateral).
- **Descrição:** Primeiro, as "Community Areas" foram divididas em dois grupos de acordo com o número de criminosos sexuais registrados com ponto de corte sendo o terceiro quartil. A partir disso, foi calculada a proporção de crimes sexuais em relação ao total de crimes em cada grupo e, então, foi aplicado o Teste 'Z' para avaliar se a proporção de crimes sexuais nas áreas de alta concentração é significativamente maior do que nas áreas de baixa concentração, por isso foi utilizado "*larger*" na variável "*alternative*".
 - **Hipóteses para Teste Z:**
 - Hipótese nula (H0): A probabilidade de ocorrer um crime sexual em áreas com alta concentração de criminosos sexuais registrados é menor à probabilidade em áreas com baixa concentração.
 - Hipótese alternativa (H1): A probabilidade de ocorrer um crime sexual em áreas com alta concentração é maior do que em áreas com baixa concentração.
- **Resultados:**
 1. Probabilidade crimes sexuais em áreas de alta concentração: 0.0096
 2. Probabilidade crimes sexuais em áreas de baixa concentração: 0.0106
 3. Teste Z para duas proporções (alta > baixa):
Estatística Z = -10.2126
p-valor = 1.0000

- **Interpretação:** Os números mostram que a relação entre crimes sexuais e o total de delitos é 0,96 % nas “Community Areas” de alta concentração de infratores sexuais, e 1,06 % nas áreas de baixa concentração. O teste Z revelou que, na realidade, a participação dos crimes sexuais é menor nas áreas de alta concentração. Em suma, a existência de muitos infratores sexuais não aumenta a chance relativa de novo crime sexual, possivelmente porque essas áreas recebem maior fiscalização ou apresentam menor densidade populacional.

5. Conclusão

A análise mostra onde e para quem concentrar esforços de prevenção criminal em Chicago. De forma geral, foi possível entender que os crimes sexuais se movem em sincronia com a criminalidade geral, mas têm taxas de resolução mais baixas. Além disso, poucos bairros concentram infratores sexuais e crimes violentos simultaneamente, sendo que apenas alguns, considerados populosos e economicamente vulneráveis, possuem volumes altos de crimes não sexuais e poucos infratores sexuais registrados. Em síntese, mapear essas discrepâncias fornece insights valiosos para decisões de patrulhamento, políticas de habitação, aumento de recursos investigativos e programas de proteção a grupos mais suscetíveis, permitindo que as autoridades maximizem o impacto das intervenções de segurança pública.

6. Referências

City of Chicago. Community Areas MAP. [S.l.], 22 abr. 2025. Disponível em: <https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Community-Areas-MAP/3fqw-rq4x>. Acesso em: 11 jun. 2025.