



Table of **Contents**

01

Intro

Background information
& Problem Statement
overview

02

Pre-Modeling

Cleaning process,
clusters & LDA!

03

EDA and Predictive Models

Review of EDA analysis
and all models created

04

Descriptive Statistics

Informing guidance on
accidental death
prevention

05

Final Thoughts

Introduction

PROBLEM STATEMENT:

- The Medical Examiner's Office has limited investigative resources allocated so accuracy and speed are paramount
- How can we establish a method to efficiently direct resources for investigating non-natural deaths in the county which have undetermined or pending manners of death?
- Can we identify commonly preventable patterns of unintentional deaths and direct interventions aimed at preventing accidents or educating the public to increase overall safety of the county

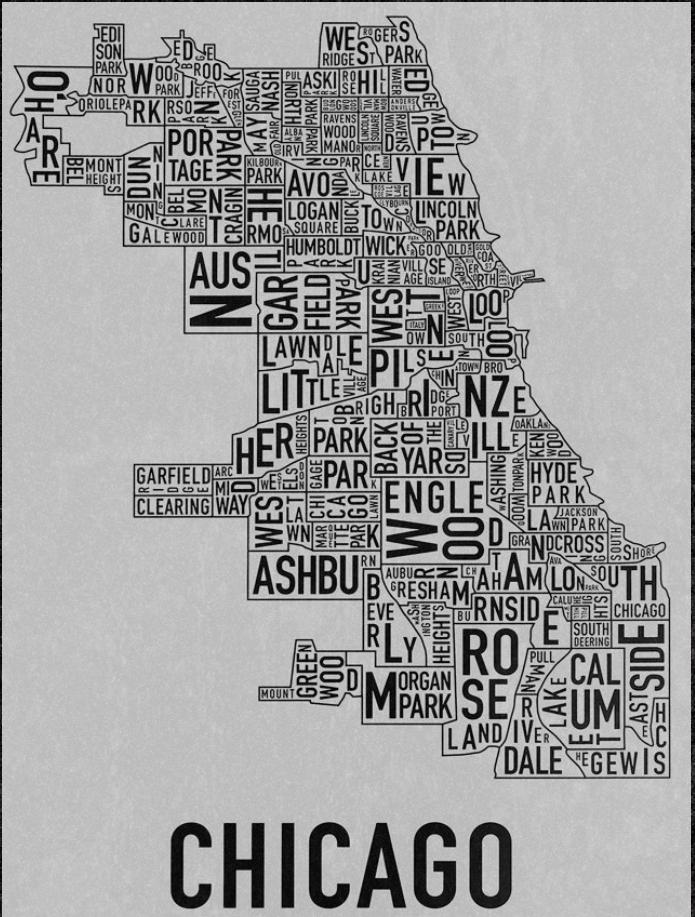


Chicago



Background and Demographics

- **Demographics:** Chicago, the third-most populous city in the U.S., boasts a diverse population mix, including Black, Hispanic, White, and Asian communities.
- **Population & Area:** With a population of around 2.7 million people, Chicago spans an area of approximately 234 square miles (606 square kilometers).
- **Cultural Hub:** Known as the "Windy City," Chicago serves as a major financial, cultural, and transportation hub in the Midwest.
- **Architectural Marvels:** Renowned for its iconic architectural landmarks like the Willis Tower and John Hancock Center, Chicago is a city that blends modernity with historic charm.



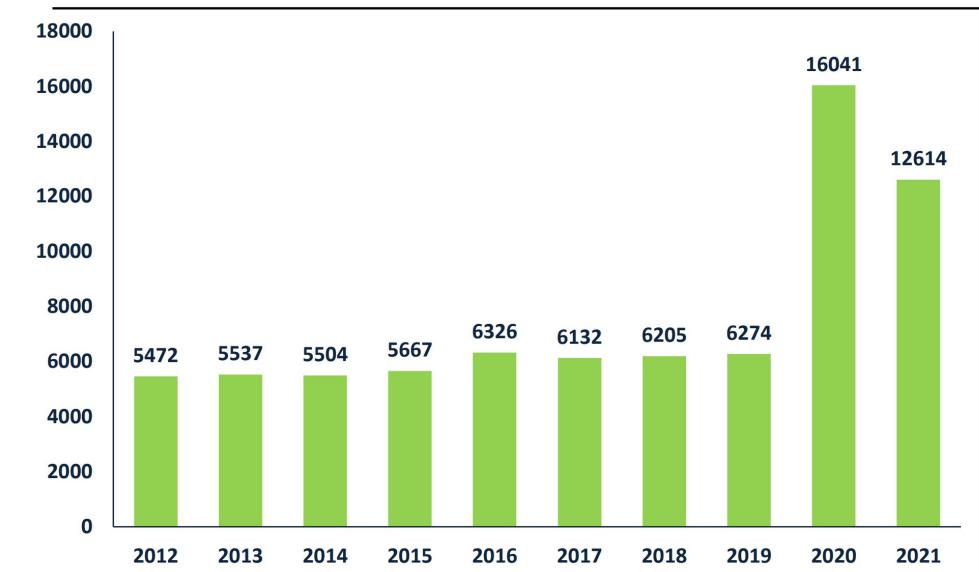
Introduction

- Year by year investigation totals
- Increased number of investigations after COVID-19

COOK COUNTY MEDICAL EXAMINER'S OFFICE | 2021 ANNUAL REPORT

COOK COUNTY MEDICAL EXAMINER CASES

Fig. 1. Number of Cases Examined by CCMEO, 2012-2021



https://www.cookcountyl.gov/sites/g/files/ywwepol61/files/documents/2023-07/2021%20CCMEO%20Annual%20Report_FINAL%20%28Web%29.pdf

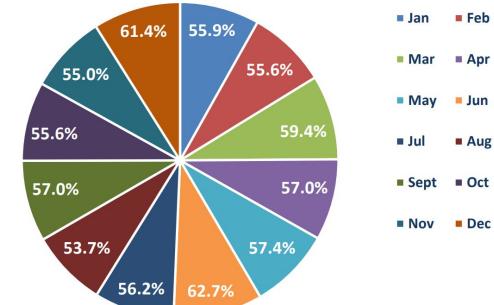
Introduction

- 2019 data
- Distribution of accepted cases and autopsies varies only slightly.
- Predictable case rate for investigators
- Reports change in appearance after this year.

Table 4. Breakdown of CCMEO Cases and Autopsies by Month, 2019.

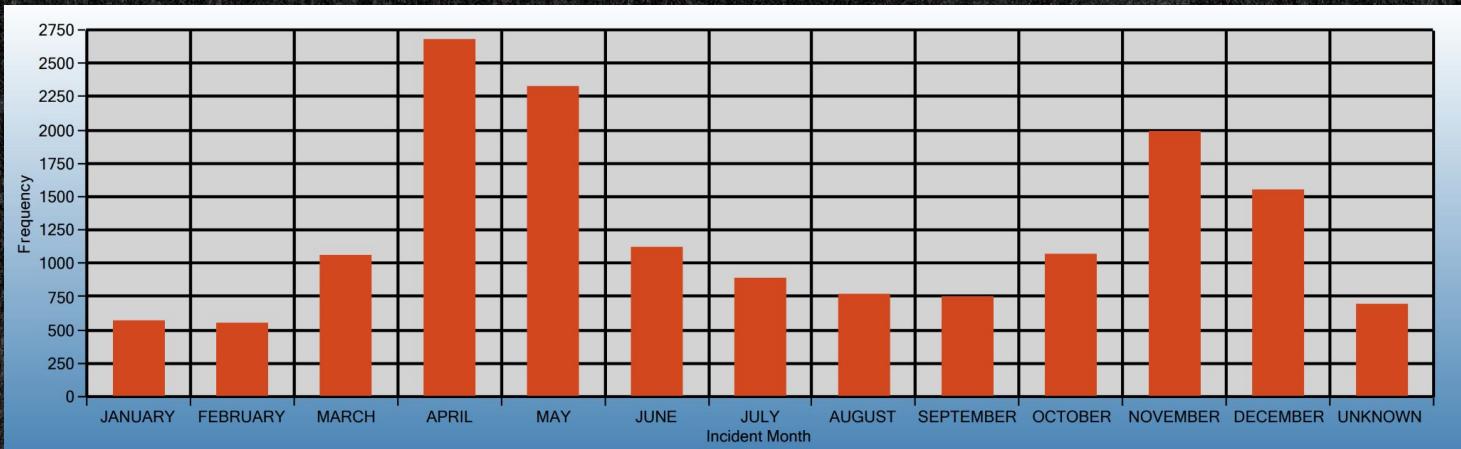
Month	Accepted Cases	Autopsies
January	495	288
February	468	271
March	466	296
April	499	276
May	622	299
June	656	307
July	620	309
August	629	273
September	555	297
October	550	285
November	513	301
December	450	358

Fig. 2: CCMEO Autopsy (%) by Month, 2019



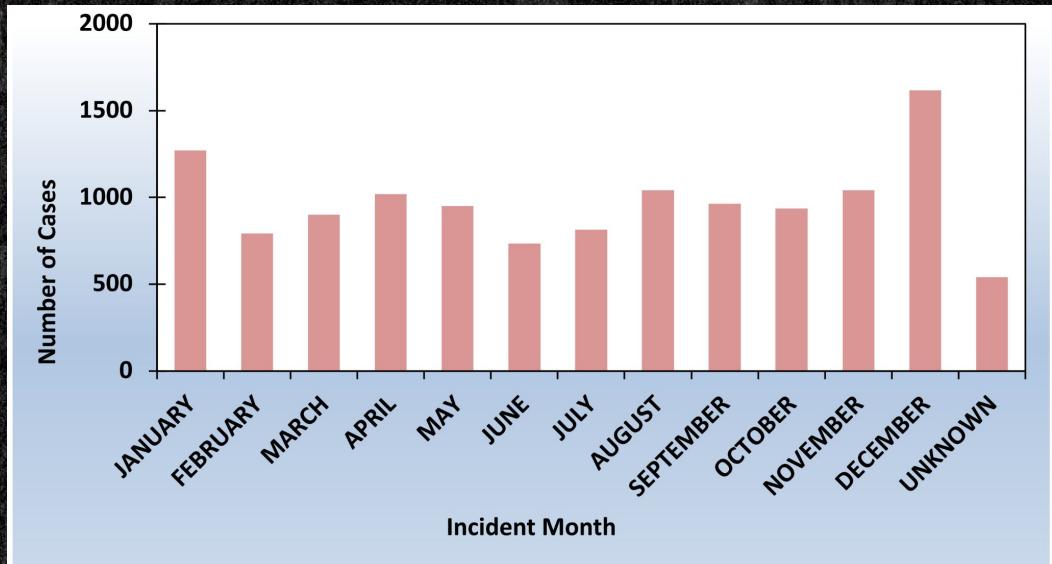
Introduction

- 2020 Monthly case breakdown
- Spike in April 2020?
- A new paradigm?



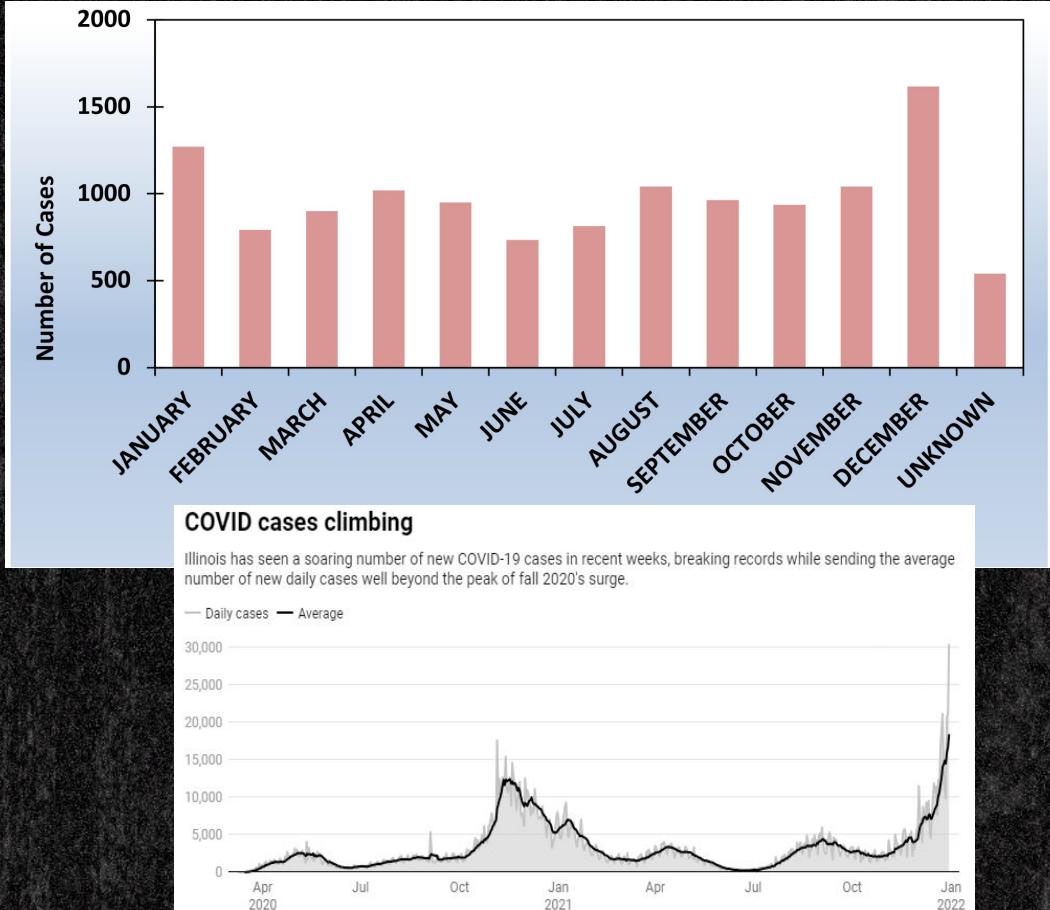
Introduction

- 2021 Monthly case breakdown



Introduction

- 2021 Monthly case breakdown
- The december 2021 spike in investigations correlations “”

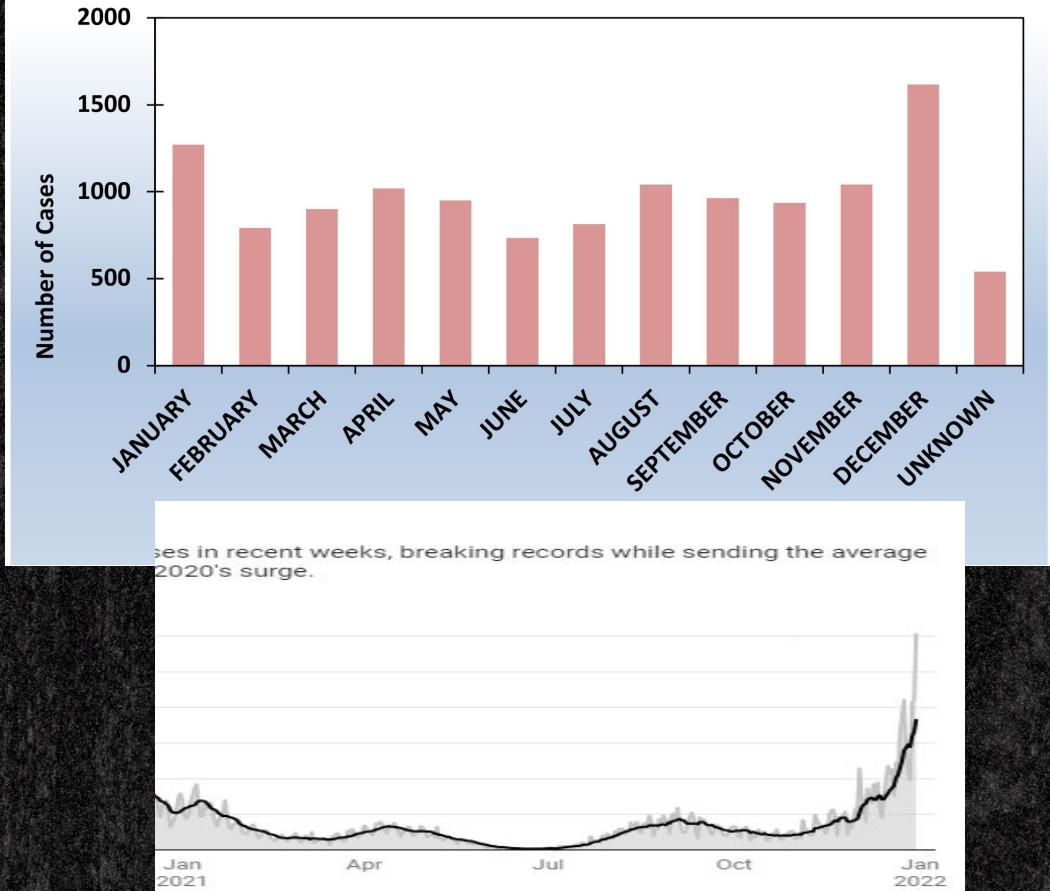


<https://www.chicagotribune.com/news/breaking/ct-covid-illinois-chicago-spike-20211228-ky7j3c4qx5agje7n5ufq6estqq-story.html>

 https://www.cookcountylil.gov/sites/g/files/ywwepo161/files/documents/2023-07/2021%20CCMEO%20Annual%20Report_FINAL%20%28Web%29.pdf

Introduction

- 2021 Monthly case breakdown
- The december 2021 spike in investigations correlations “ ”

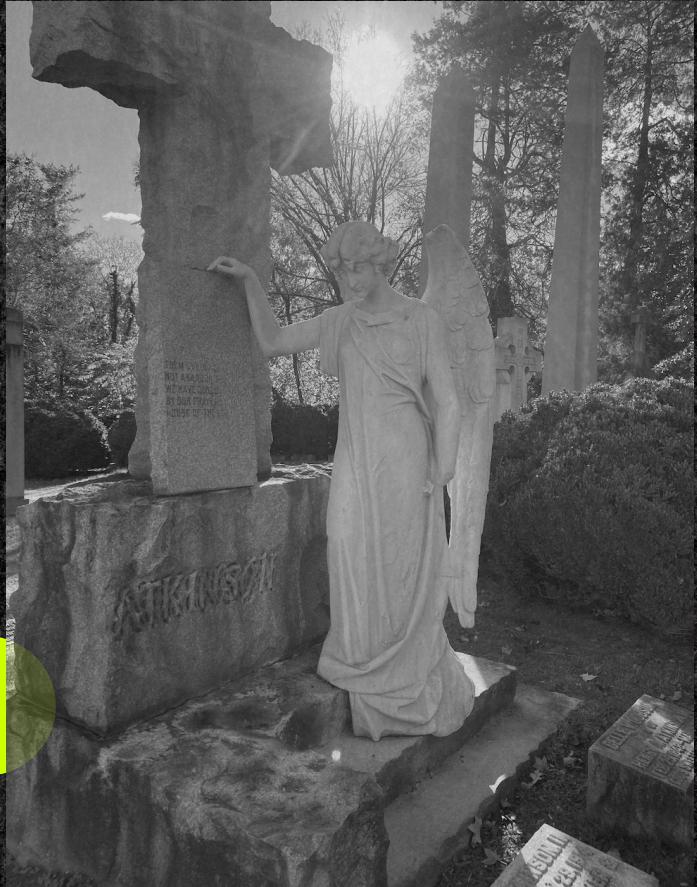


Pre-Modeling: Data Cleaning

- Many Columns were dropped as they were not very relevant
 - Case Number
 - Incident Address
 - Location
 - ObjectID
 - Chicago Community Ward
- Many Columns were added to simplify things
 - Age range
 - Death Date, Time and Day
 - Incident Date, Time and Date
- Empty Data boxes were filled with NaN, or 0 or dropped accordingly
- True/False or Yes/No values were changed into 0 or 1

Exploratory Data Analysis

Understanding the dataset



Age Distribution



Biological Gender

- 27,997 Male
 - 9,154 Female
-
- 75% Male
 - 25% Female

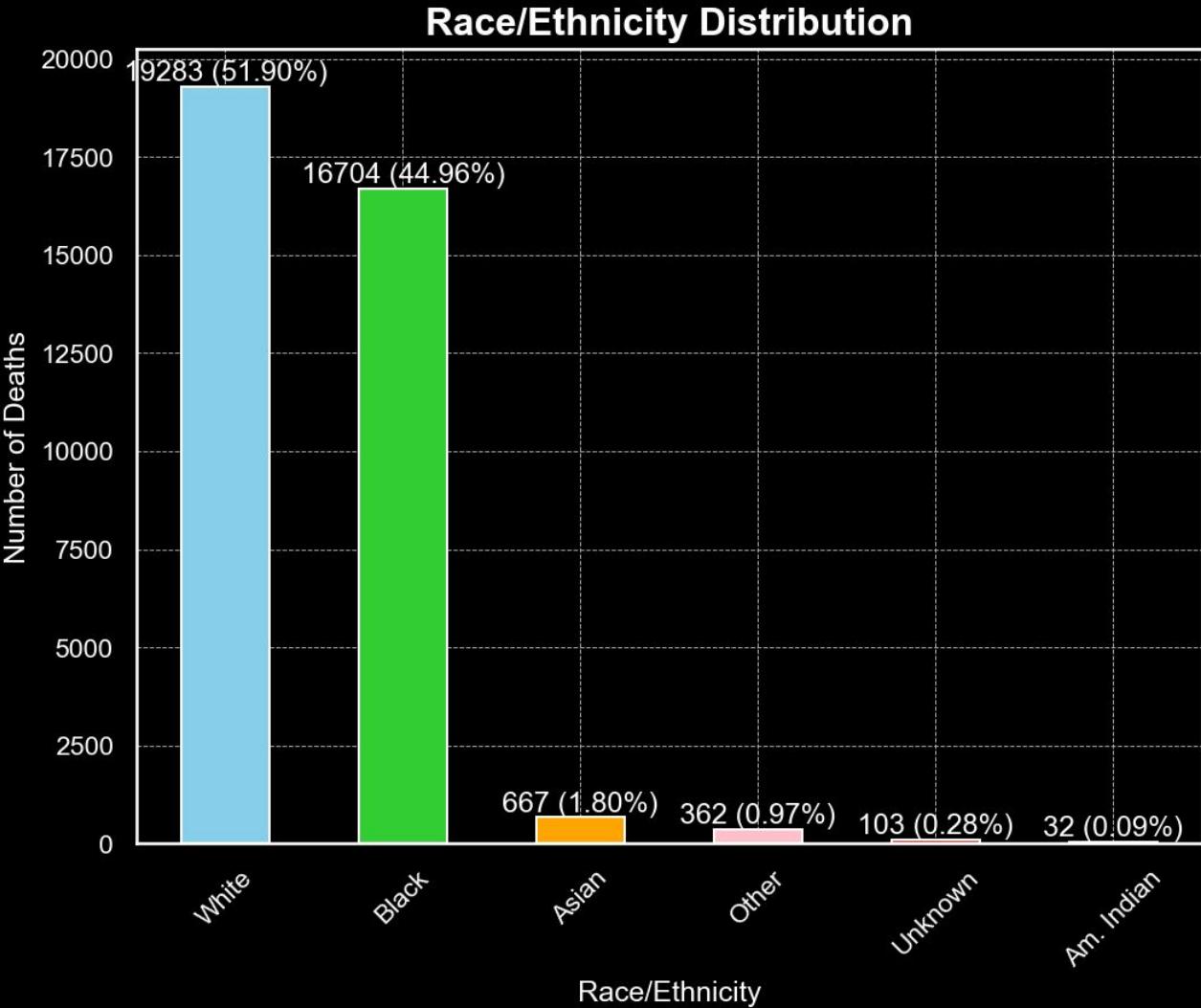
Gender Distribution



Race and Ethnicity

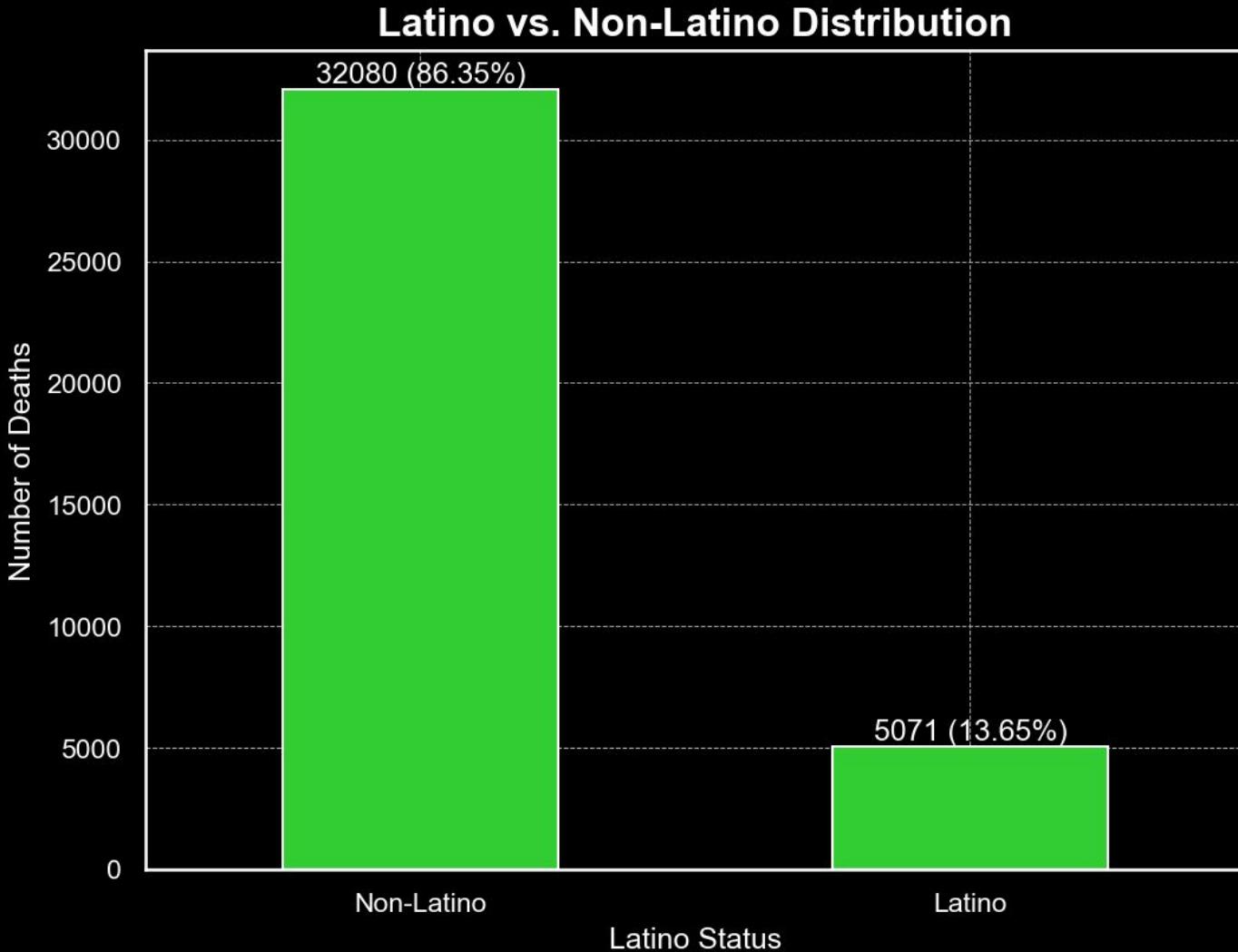
Based on Chicago census data:

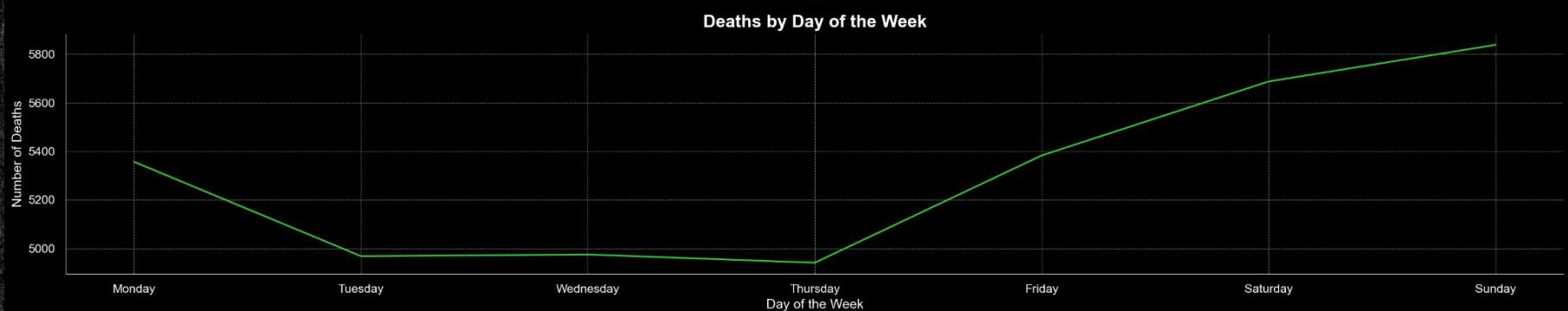
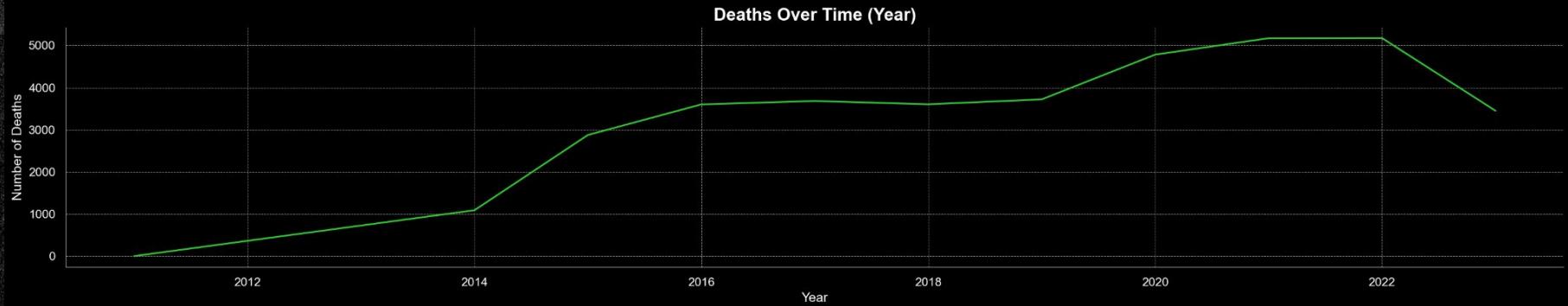
- White: 45.3%
- Black or African American: 29.2%



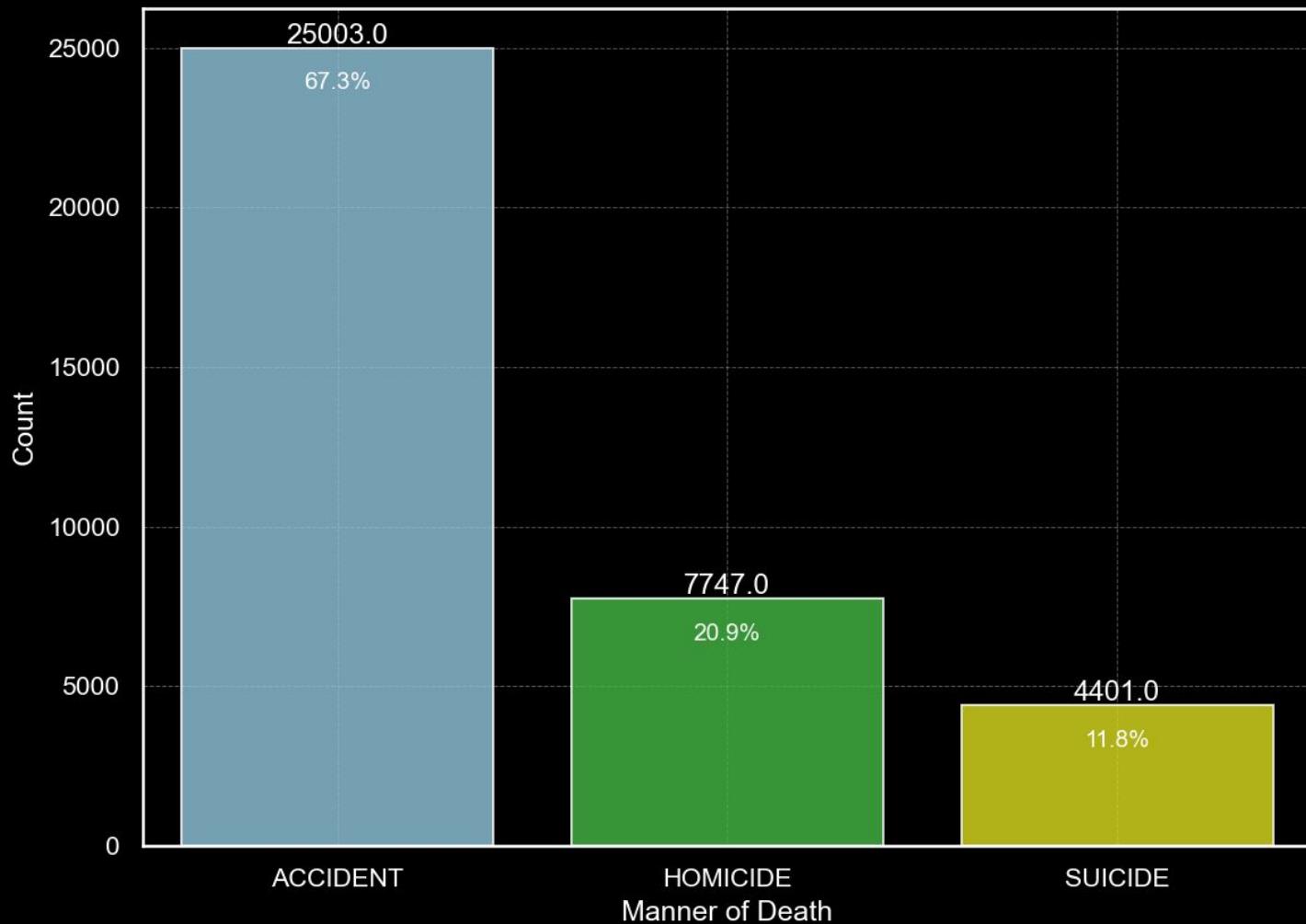
Race and Ethnicity

- White: 19283
- Black: 16704
- Latino: 5071
- Asian: 667
- Other: 362
- Unknown: 103
- American Indian: 32

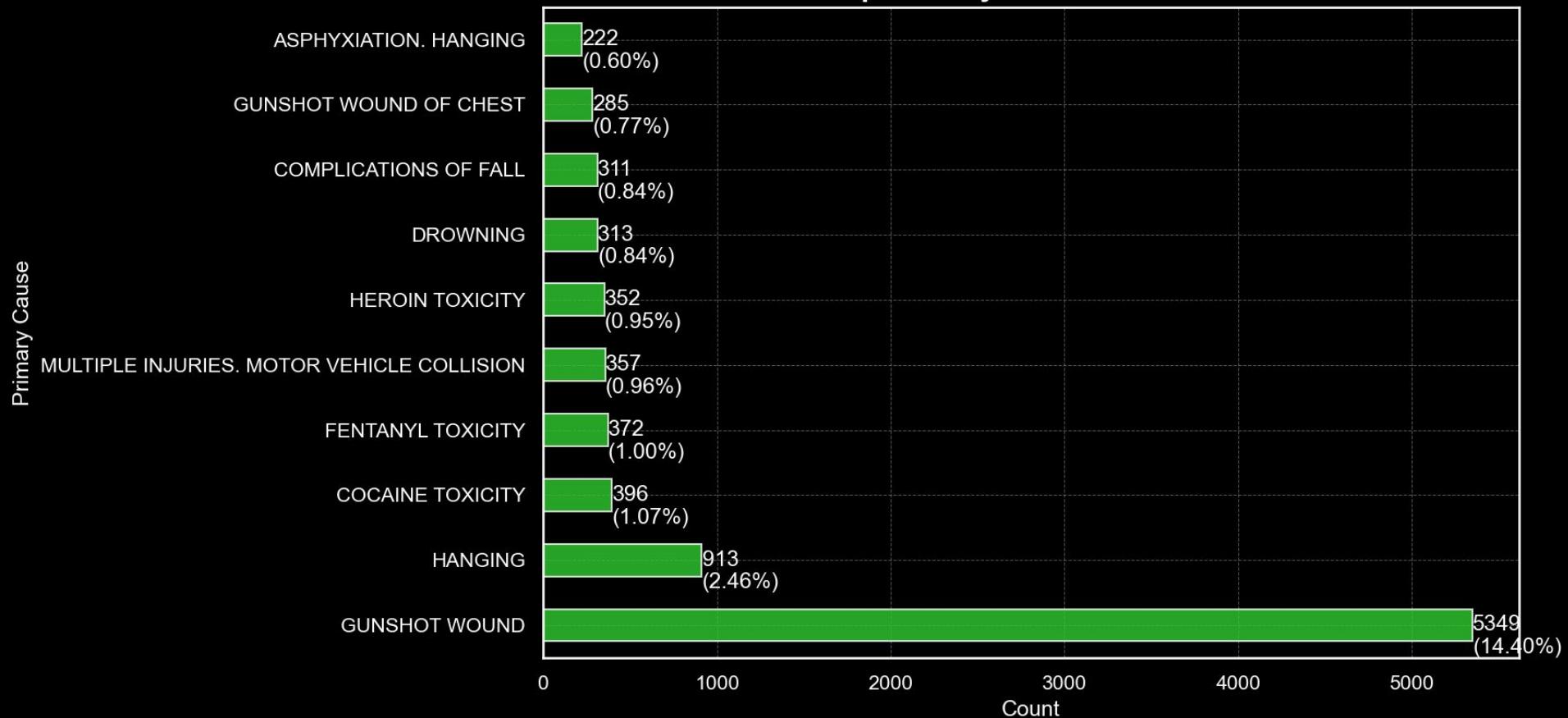




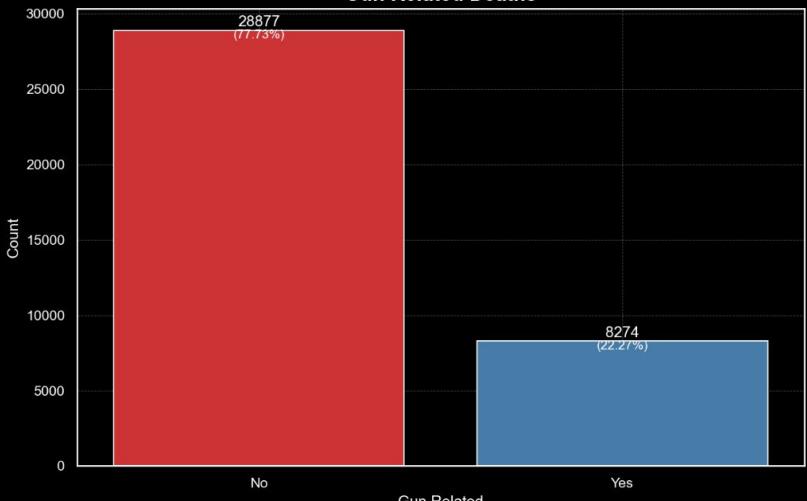
Manner of Death



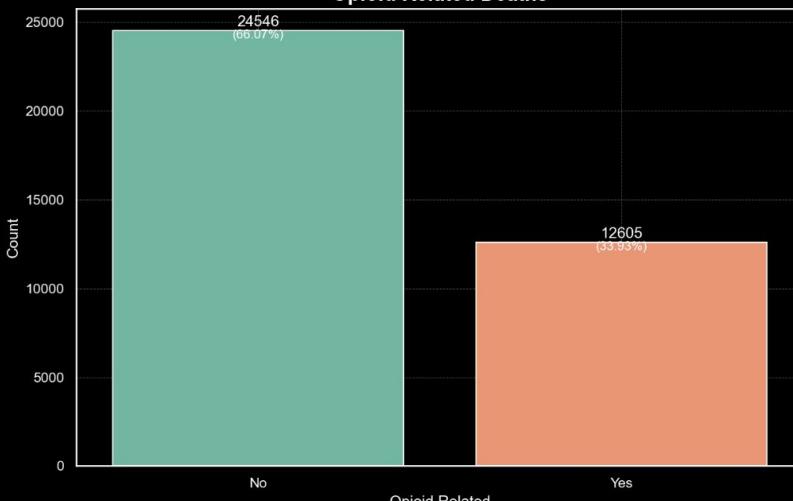
Top Primary Causes of Death



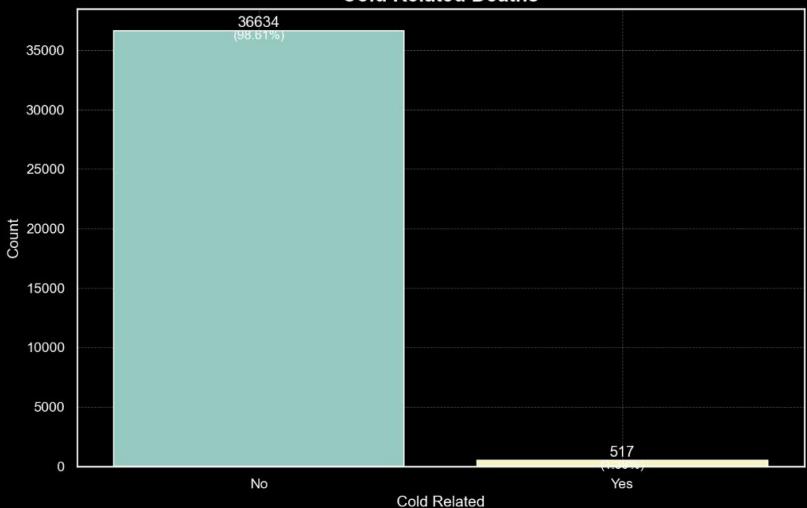
Gun Related Deaths



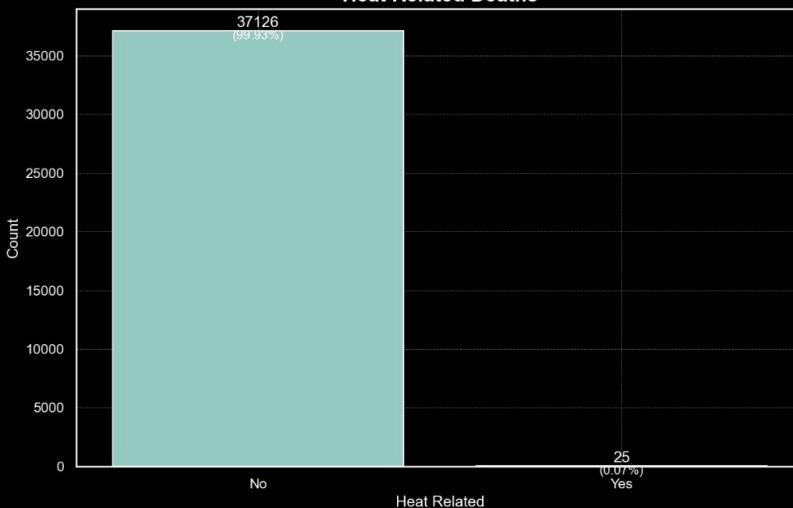
Opioid Related Deaths



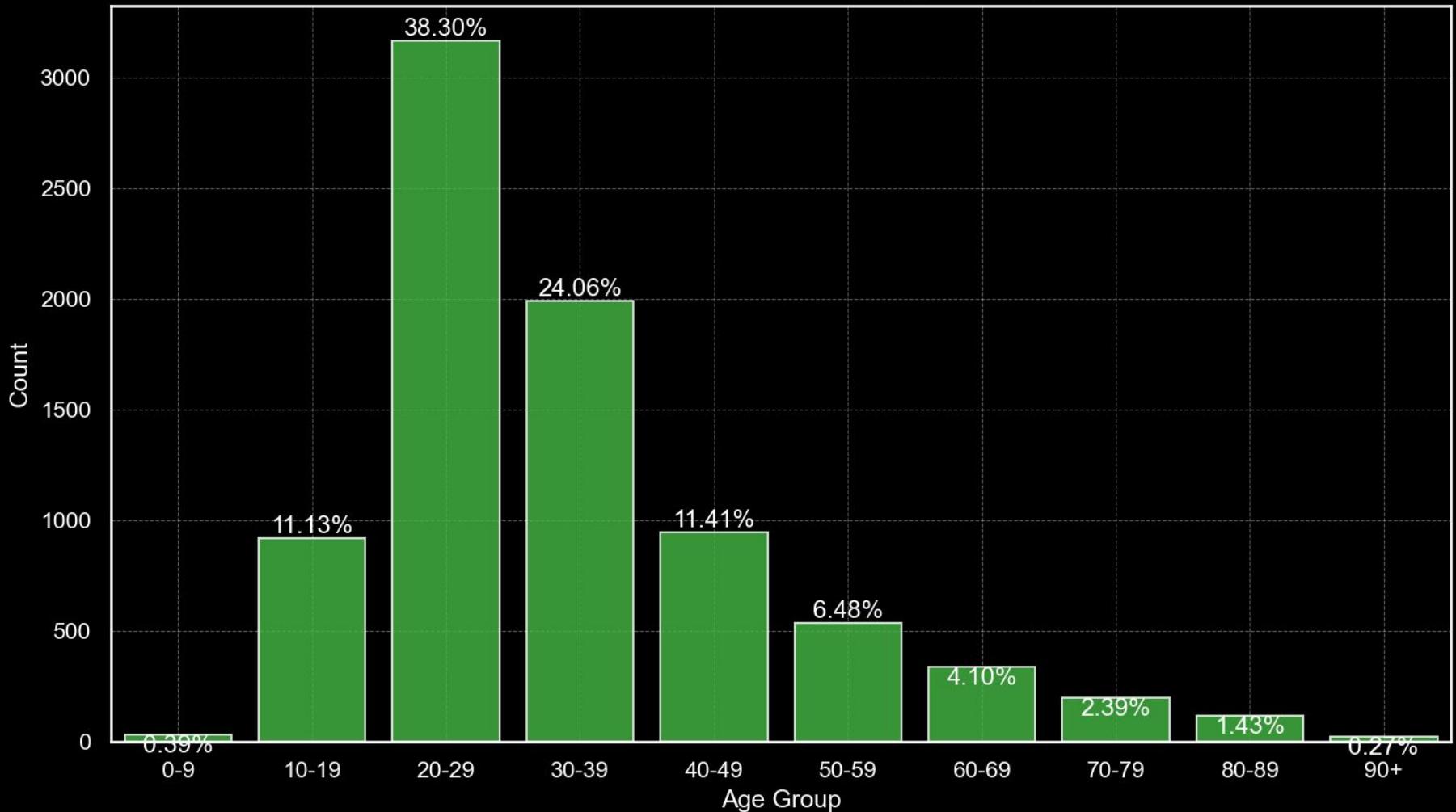
Cold Related Deaths



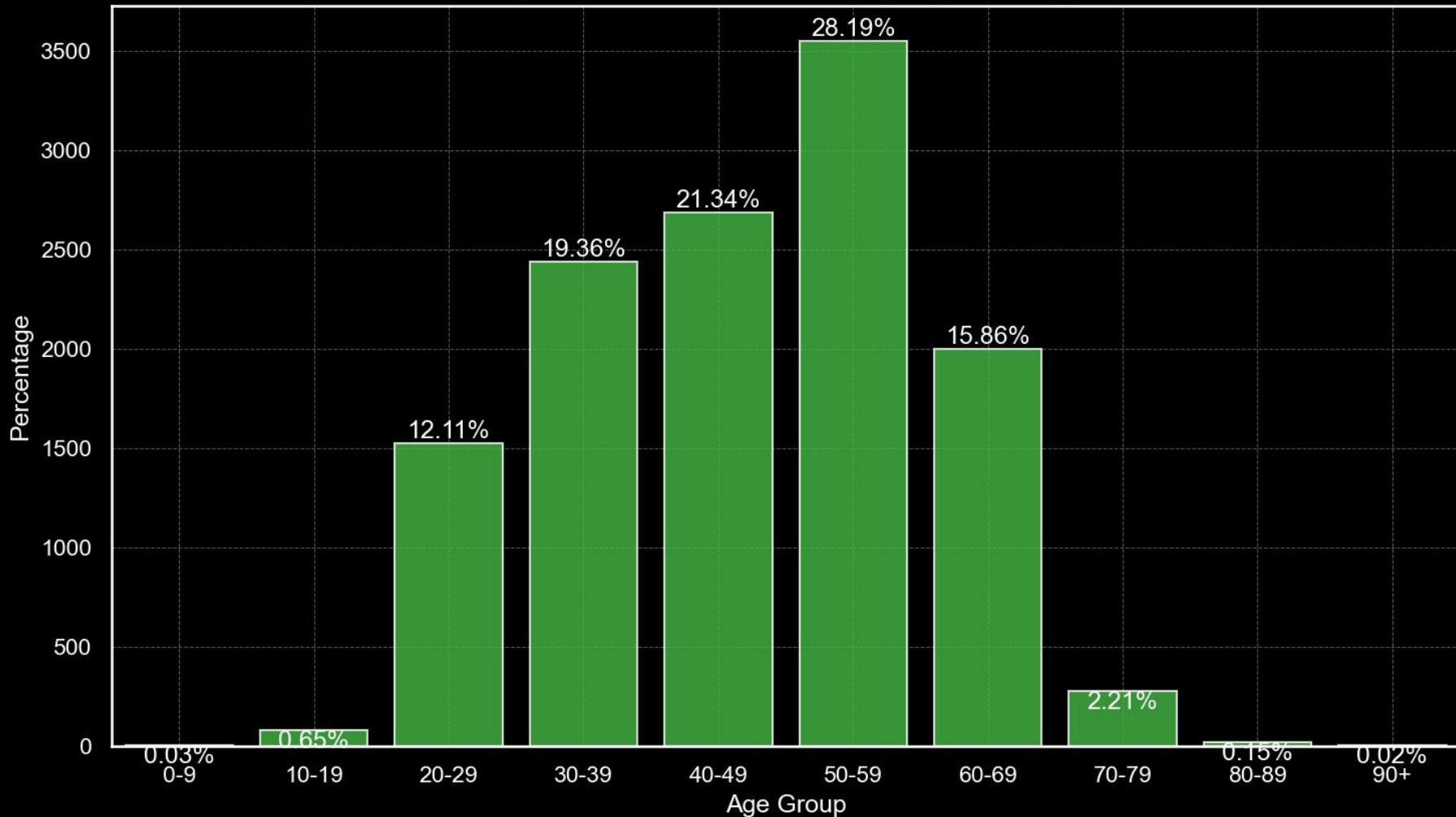
Heat Related Deaths



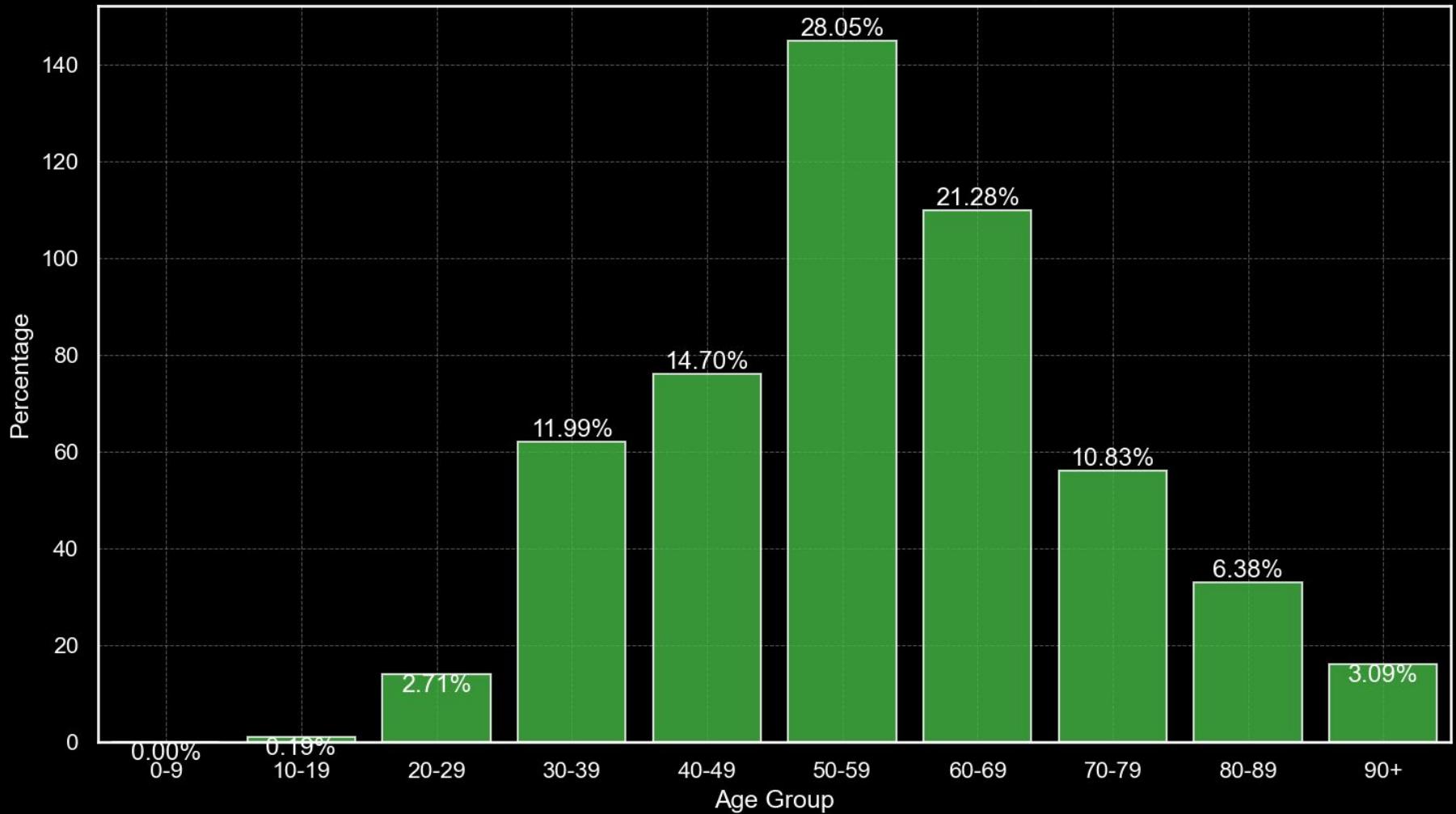
Gun Related Deaths by Age Group



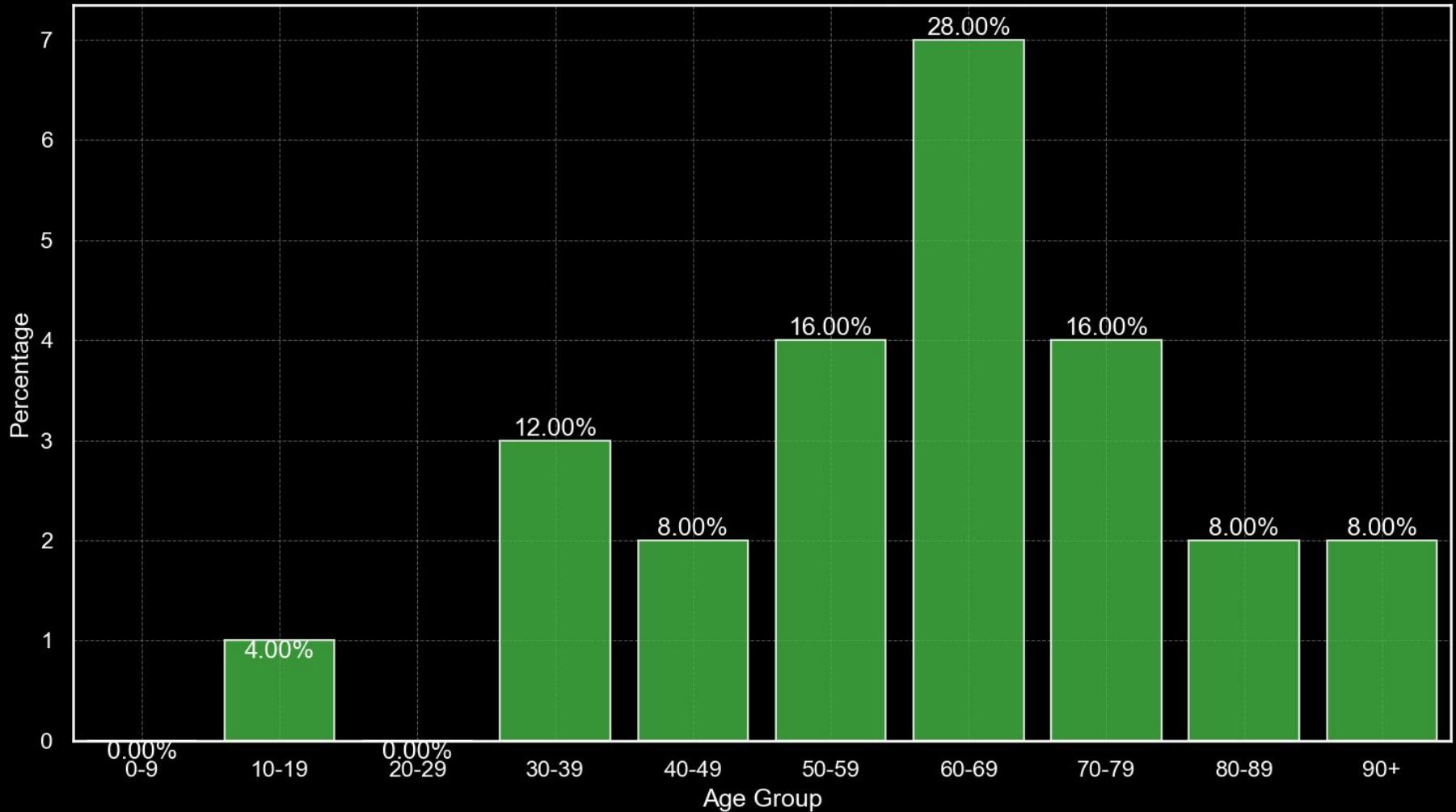
Opioid Related Deaths by Age Group



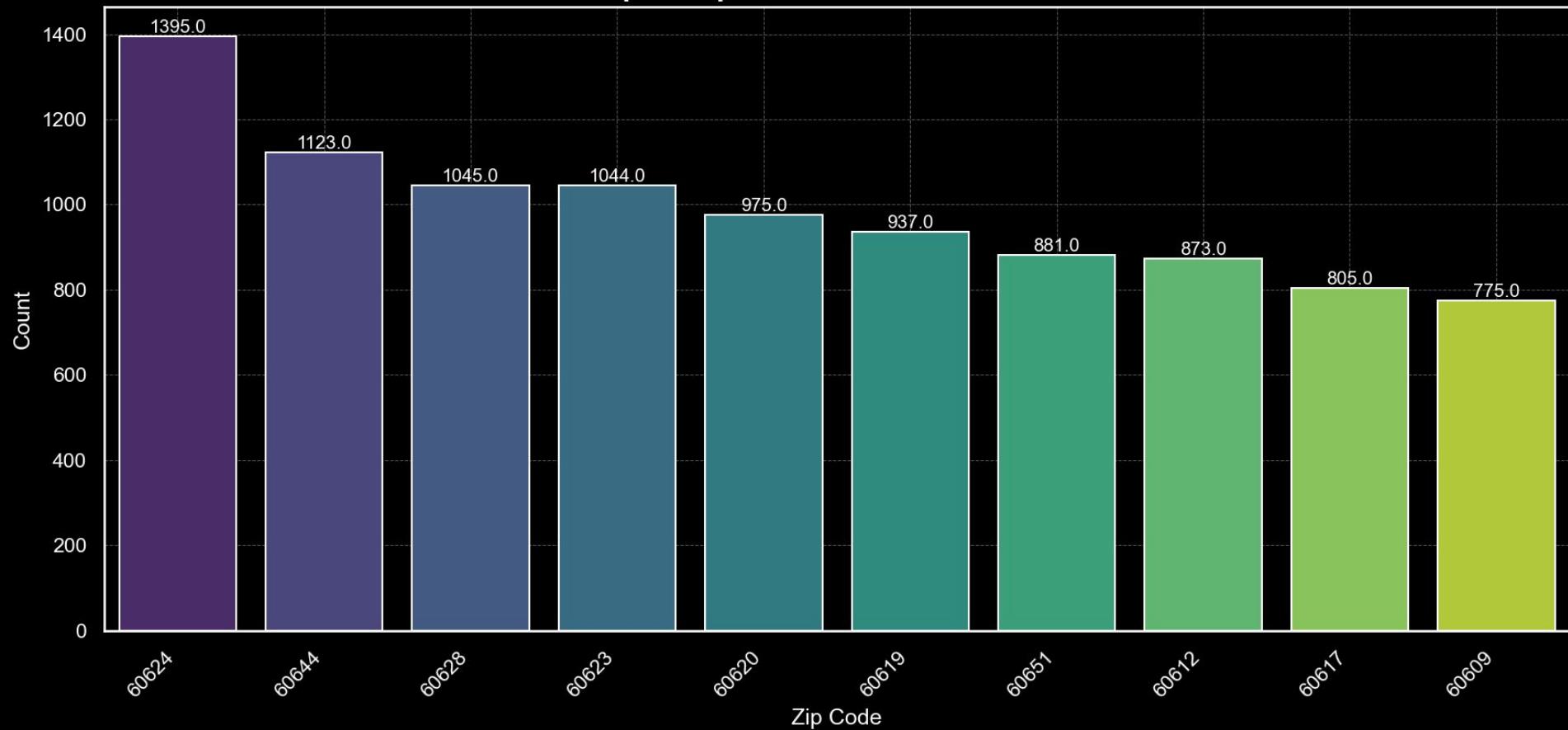
Cold Related Deaths by Age Group



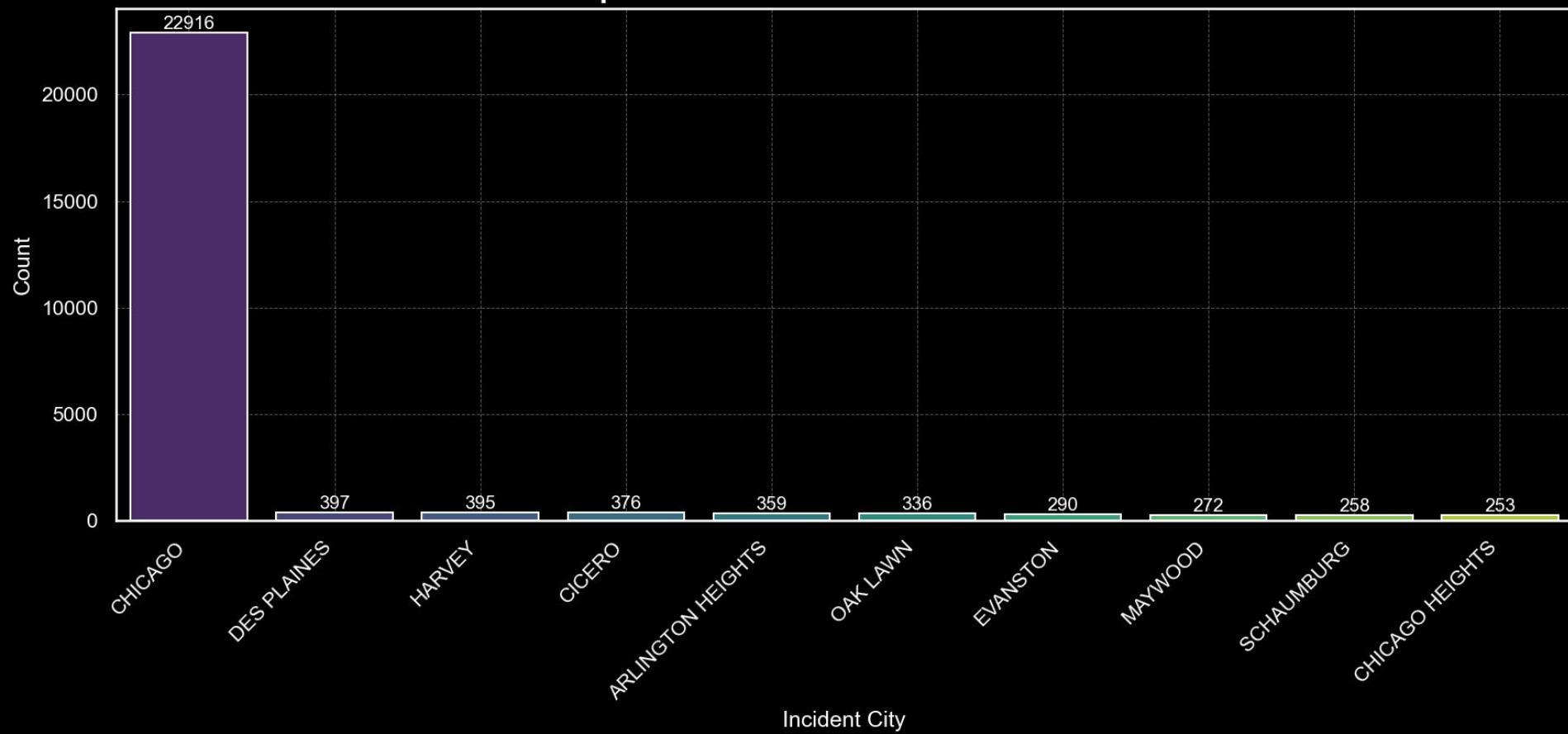
Heat Related Deaths by Age Group



Top 10 Zip Codes with Most Deaths



Top 10 Incident Cities with Most Deaths





Pre-Modeling:

Latent Dirichlet Allocation (LDA)

Topic Modeling

Topic modeling:

- Unsupervised classification of documents

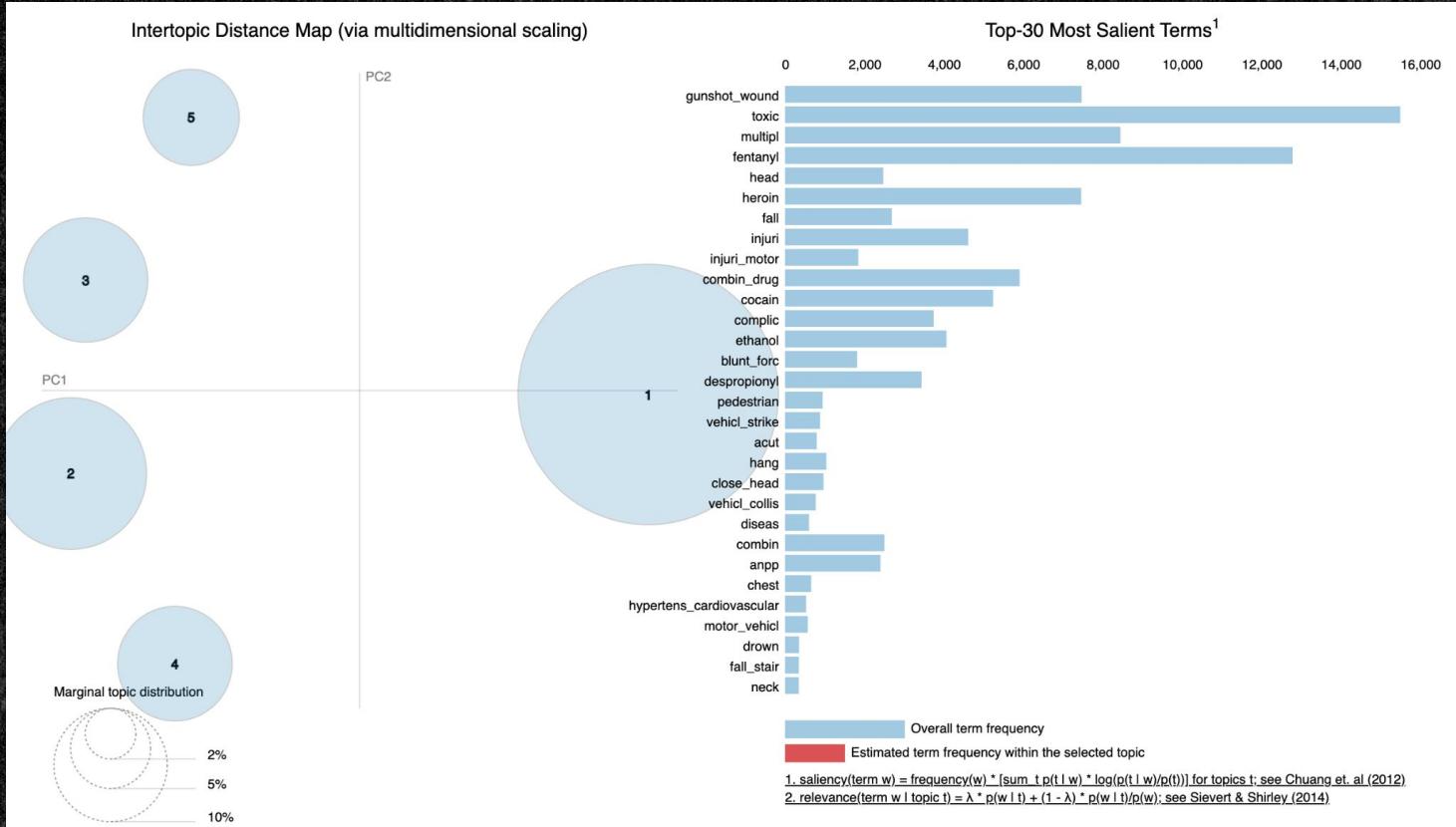
LDA:

- Each document is a bag of some words
- Each topic is a bag of some words
- Calculate proportion of words in document d that are assigned to topic t
- Calculate proportion of assignments to topic t that are because of word w
- Multiply these two to get the probability that word w belongs in topic t
- Iterate
- Assess the topic a document belongs to based on its words' topics and probabilities

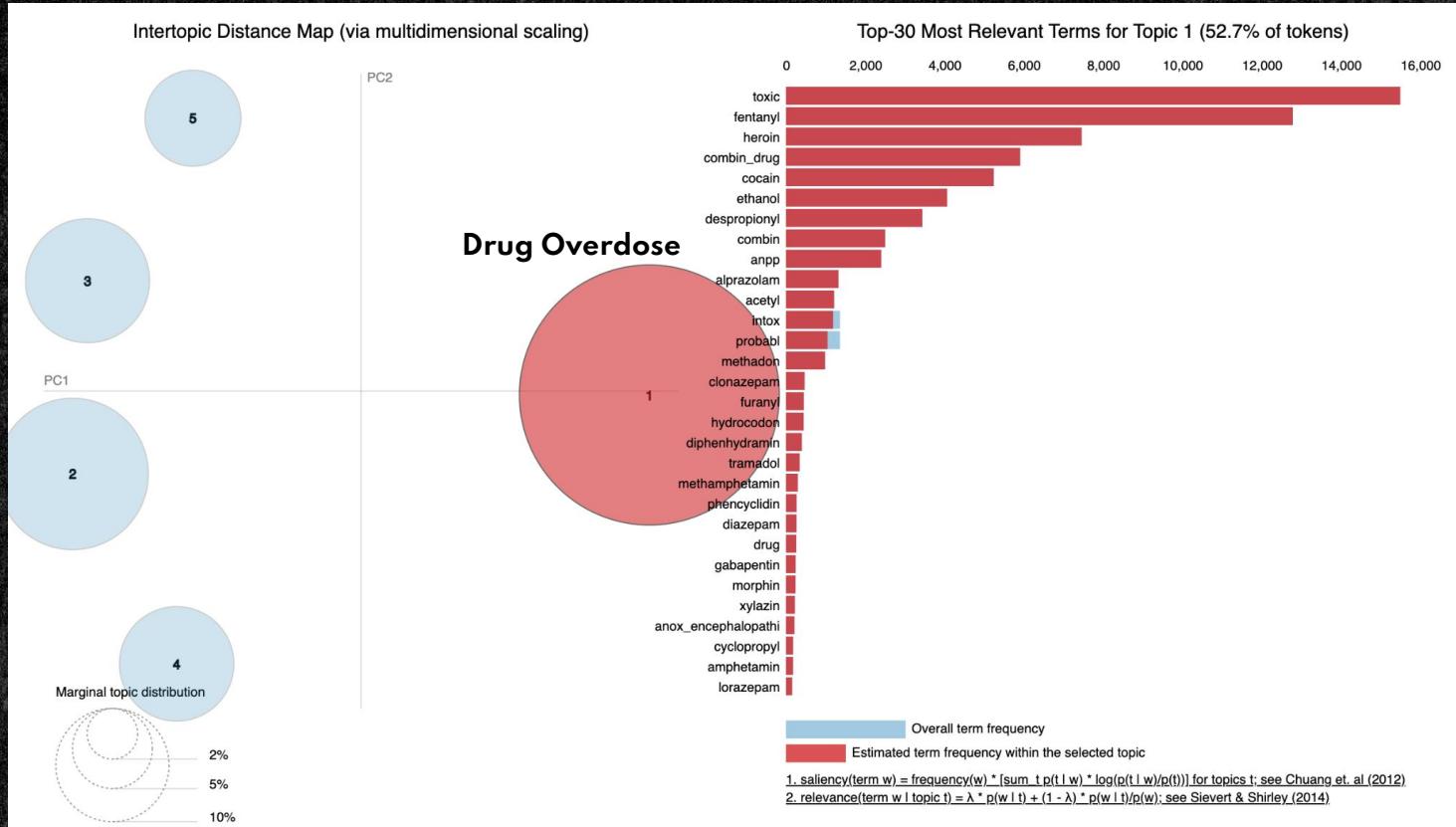
Credit: Ria Kulshrestha,

<https://towardsdatascience.com/latent-dirichlet-allocation-lda-9d1cd064ffa2>

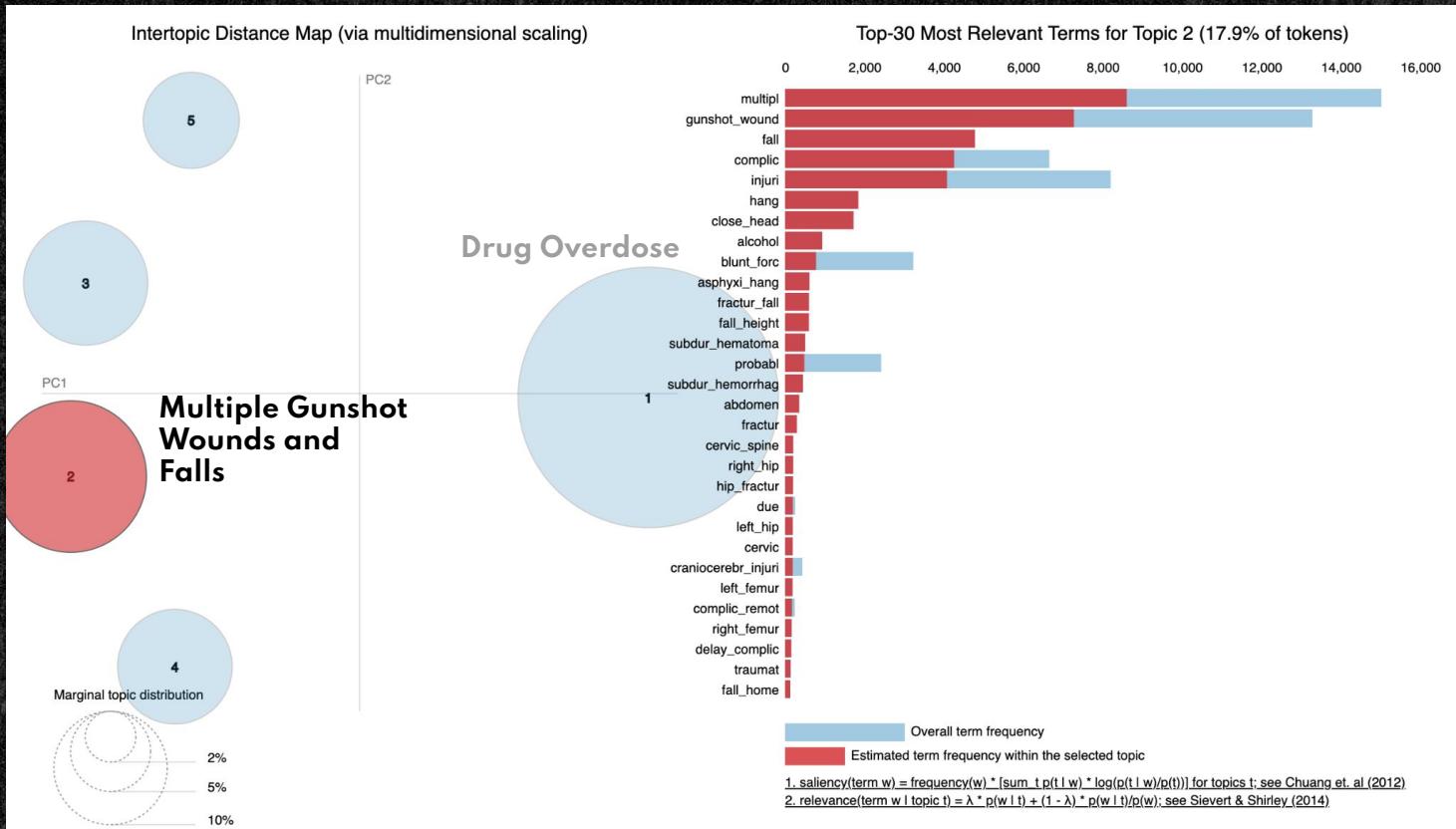
LDA Topics



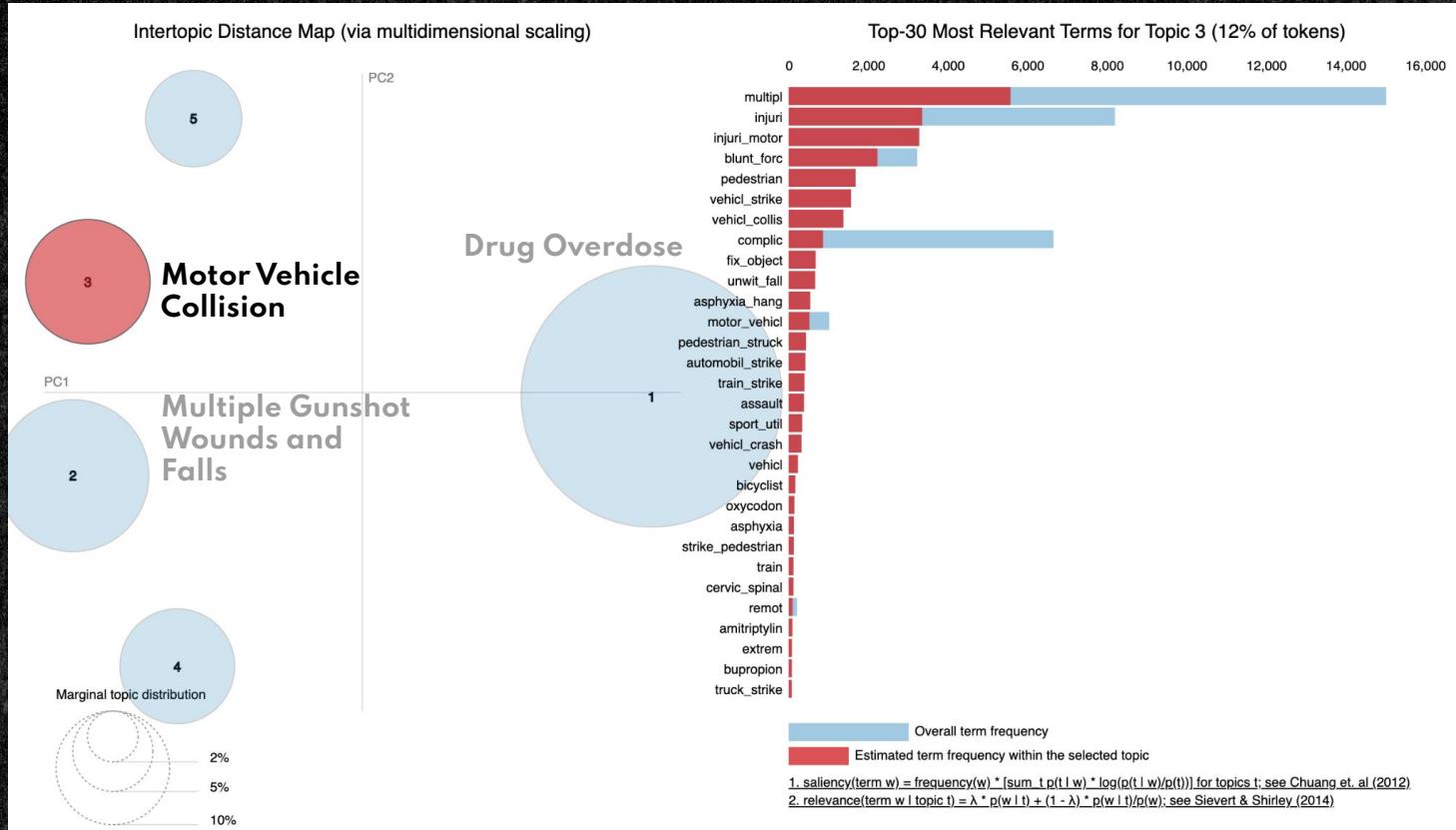
LDA Topics



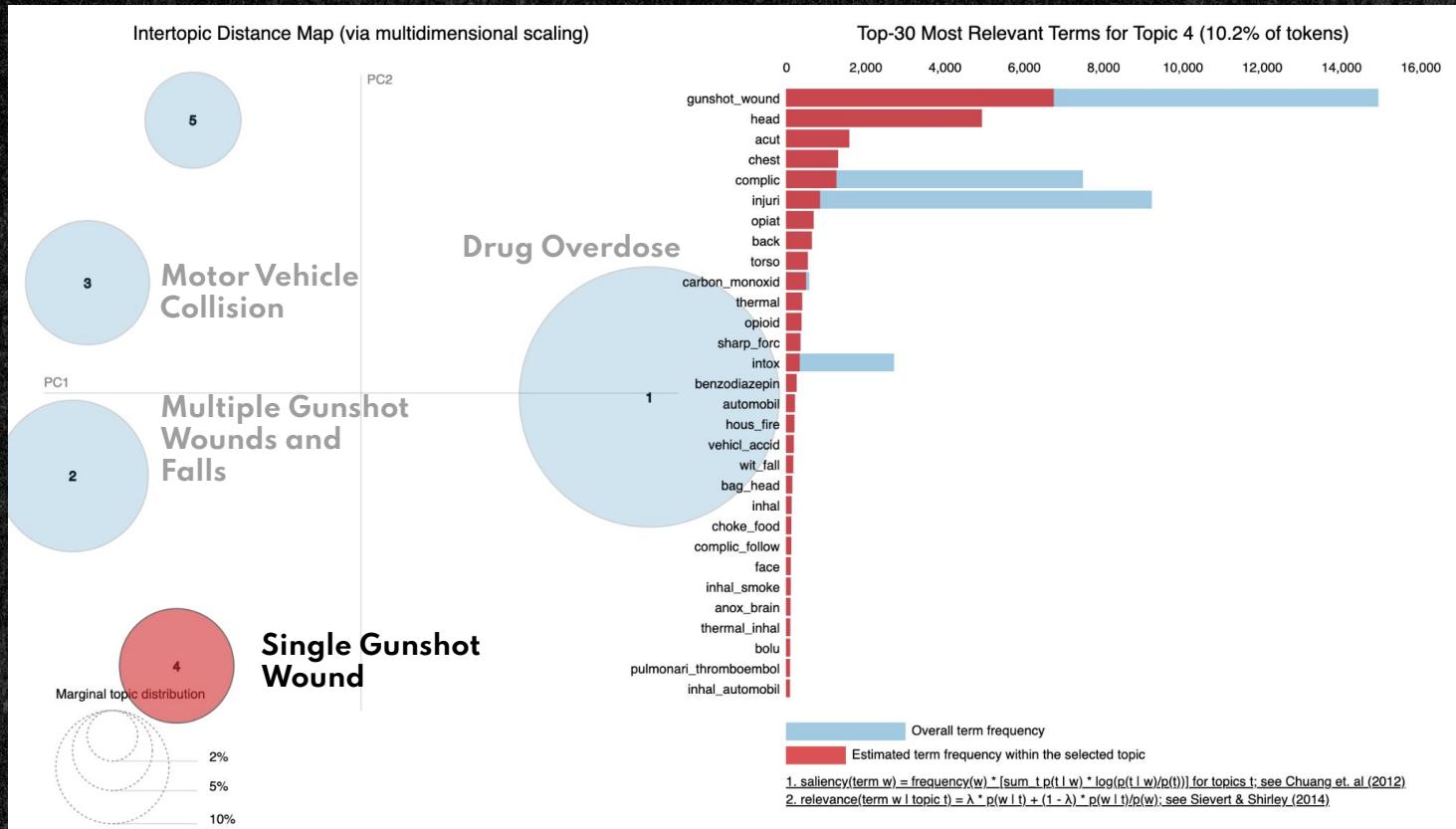
LDA Topics



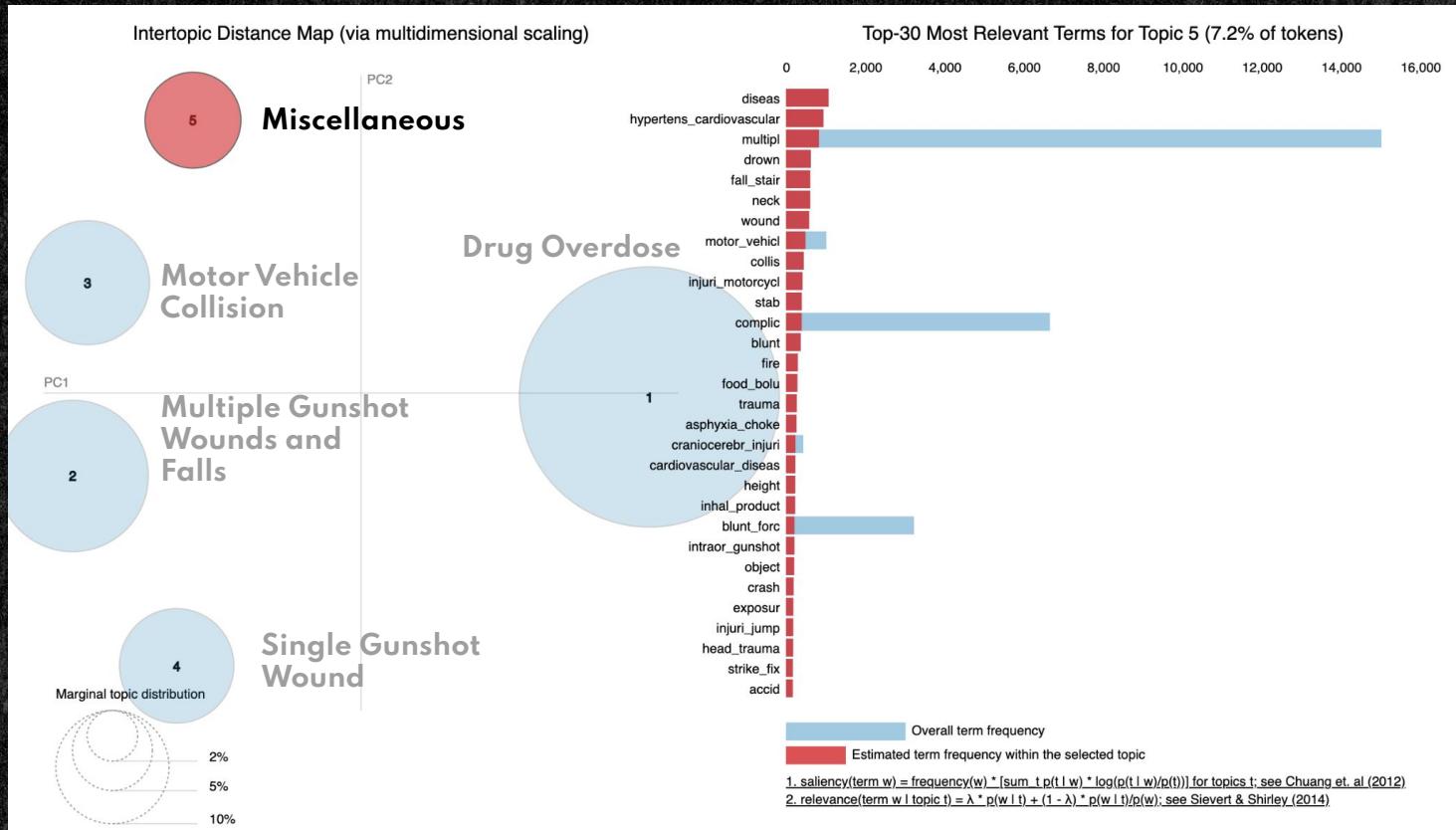
LDA Topics



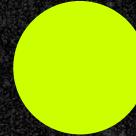
LDA Topics



LDA Topics



Models NLP



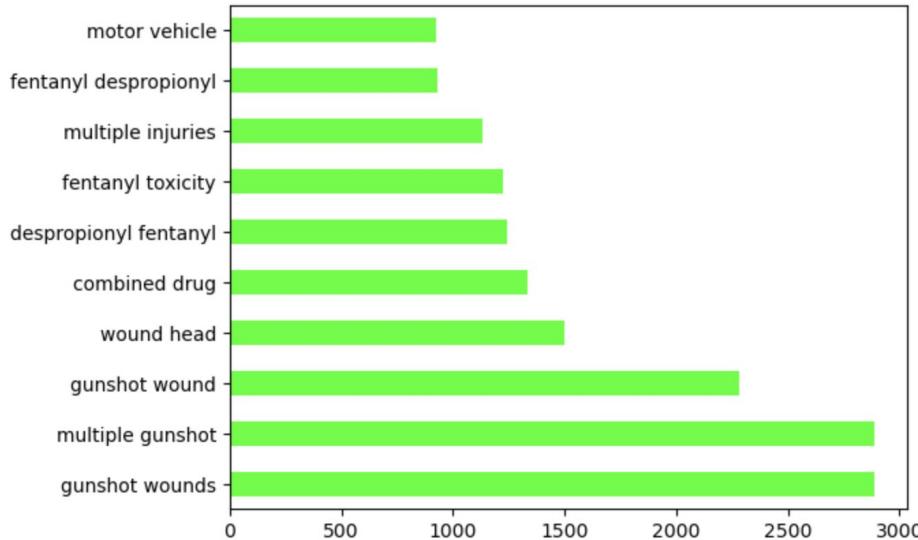
Text Extractor	Model Type	Accuracy
CountVectorizer	Multinomial Naive Bayes	97.13%
CountVectorizer	LogRegression	97.58%
TfidfVectorizer	Multinomial Naive Bayes	97.42%
TfidfVectorizer	LogRegression	97.42%

Models NLP

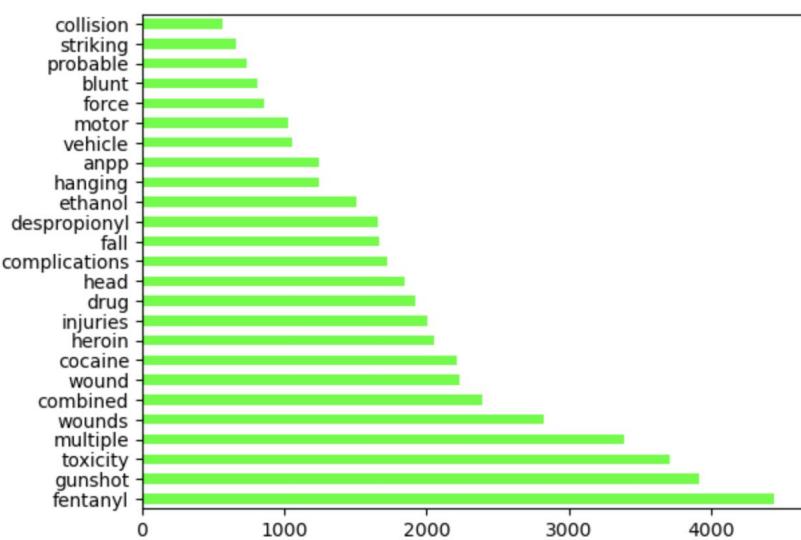
Text Extractor	Model Type	Accuracy
CountVectorizer	LogRegression	97.71%



Top 10 Primary Cause n-grams



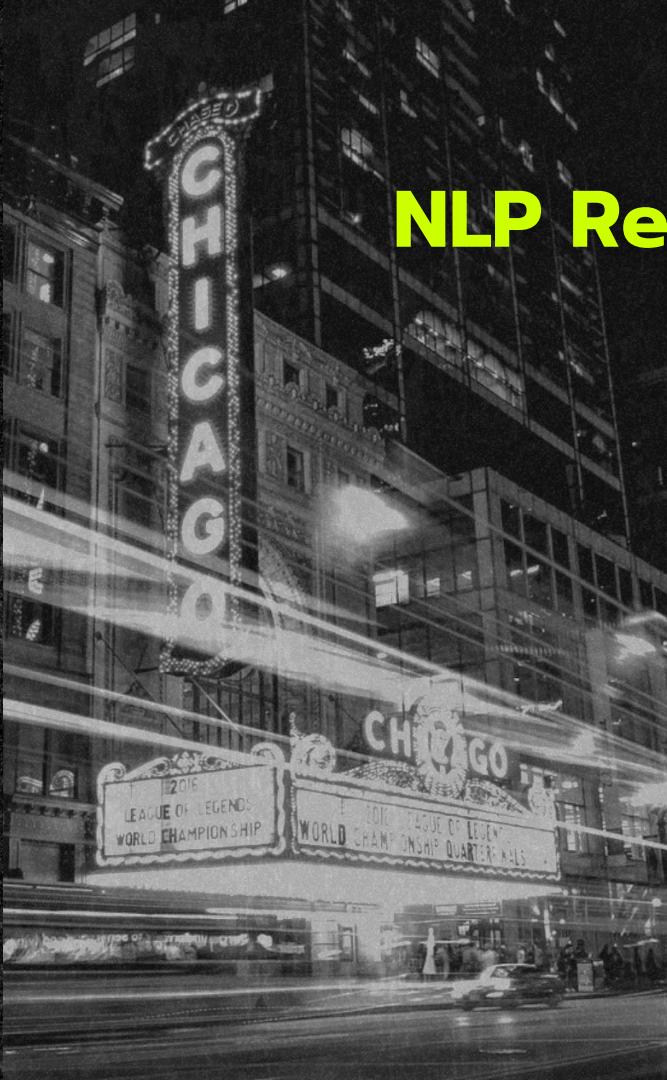
Top 25 Primary Cause Single Words



Models #1 - The Gridsearch of Gridsearch

Date/Time/Location Included?	Model Type	Accuracy
No	LogRegression	89.83%
No	KNN	89.83%
No	Decision Tree	90.62%
No	Bagged Decision tree	90.70%
No	Random Forest	90.69%
No	Ada Boost	89.92%

Date/Time/Location Included?	Model Type	Accuracy
Yes	LogRegression	82.20%
Yes	KNN	62.38%
Yes	Decision Tree	89.50%
Yes	Bagged Decision tree	90.41%
Yes	Random Forest	90.93%
Yes	Ada Boost	89.91%



Model #2

NLP Recurrent Neural Network

- Bidirectional LSTM
- Using Pre-trained glove
- Test Accuracy: 97%



Reasons for Using a Pre-trained RNN Model

- We believe that the order of the words DO matter, and the RNN model takes the words' sequence into consideration
- The words in the 'primary cause' columns of our data are brief, repetitive, and follow certain conventions, which we believe is not a good source to build a self-training model



RNN Performance on Testing Data

	Predicted Non-Accidental	Predicted Accidental
Actual Non-Accidental	2803	142
Actual Accidental	88	6139

Precision: **97.9%** Recall: **98.6%** Baseline: **67.9%**



Discussion of Incorrect Predictions



Primary Cause	Actual Manner of Death	Predicted Manner of Death	Confidence
COMBINED DRUG (TRAMADOL, GABAPENTIN, AND MORPHINE) TOXICITY	SUICIDE	ACCIDENT	62%
COMBINED ESZOPICLONE/ZOPICLONE, ZOLPIDEN, TRAZODONE, AND ETHANOL TOXICITY	SUICIDE	ACCIDENT	90%
DROWNING. NEGLECT IN BATHTUB	HOMICIDE	ACCIDENT	90%

Secondary Goal: Reducing Accidental Deaths

Category	Source
Opioid-Related	Cook County ME's Office
Gun-Related	Cook County ME's Office
Heat-Related	Cook County ME's Office
Cold-Related	Cook County ME's Office
Vehicle Crash-Related	Topic Modeling
Fall-Related	Topic Modeling

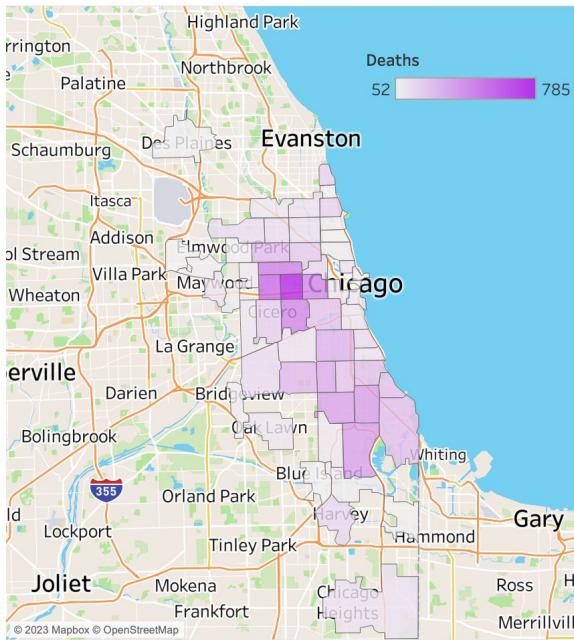


Reducing Opioid-Related Accidental Deaths

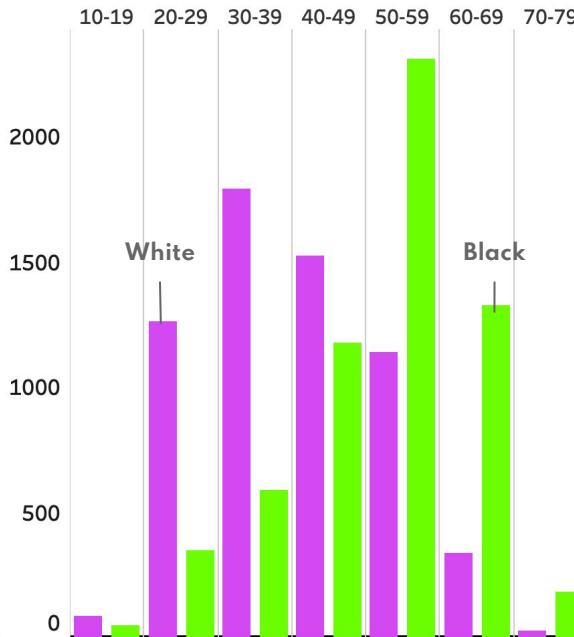


Number of Deaths Per:

Zip Code of Incident Causing Death



Race and Age Group



Recommendation:

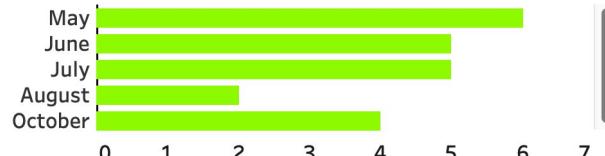
Increase awareness and accessibility of Narcan in key neighborhoods, especially among peers of :

- 20 to 50-year-old White Chicagoans
- 40 to 70-year-old Black Chicagoans

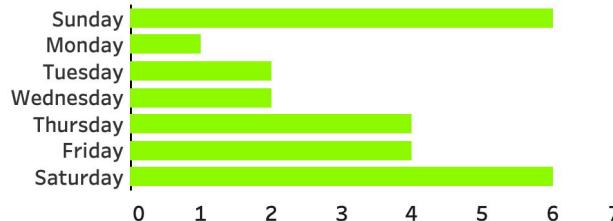
Reducing Heat-Related Accidental Deaths

Number of Deaths Per:

Month When Incident Causing Death Occurred



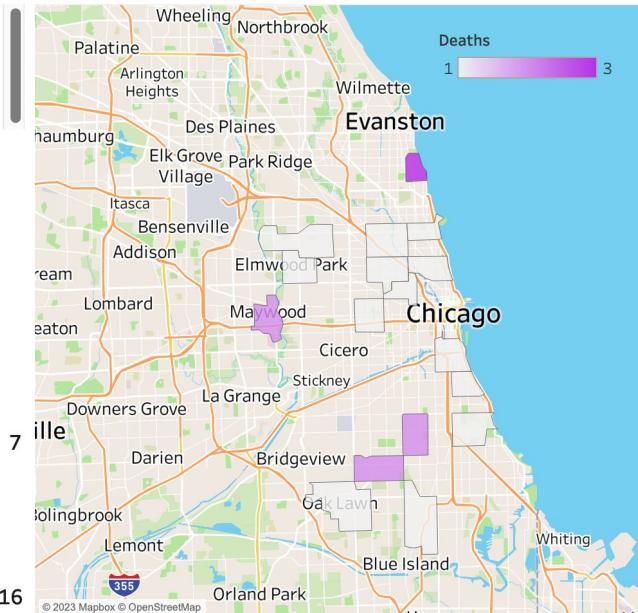
Day When Incident Causing Death Occurred



Sex



Zip Code Where Incident Causing Death Occurred



Recommendations:

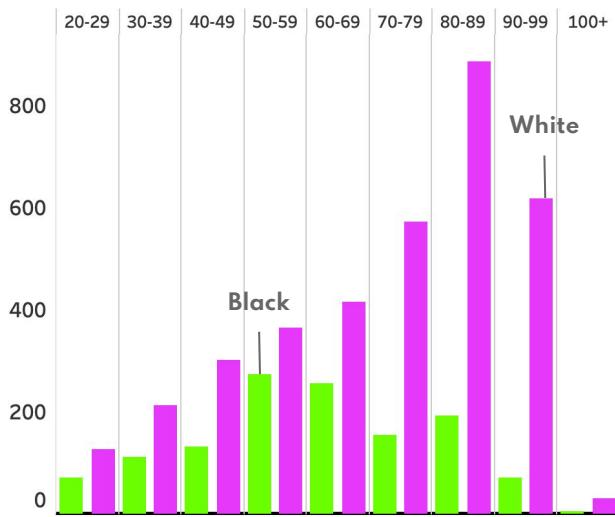
- Extend months and days of cooling center operations
- Increase awareness and transport options

Reducing Fall-Related Accidental Deaths

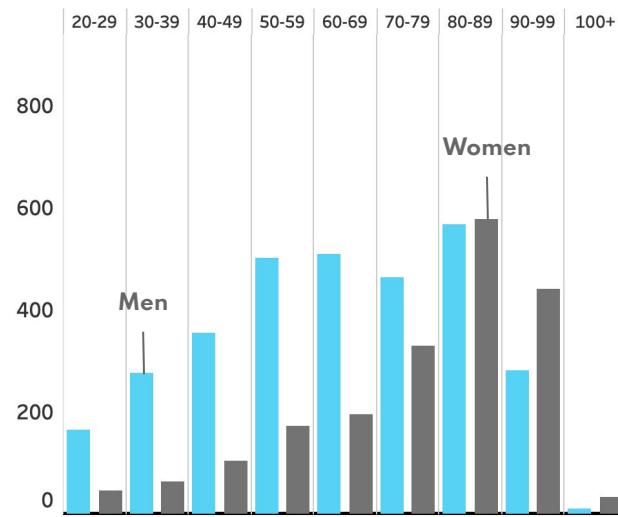


Number of Deaths Per:

Race and Age Group



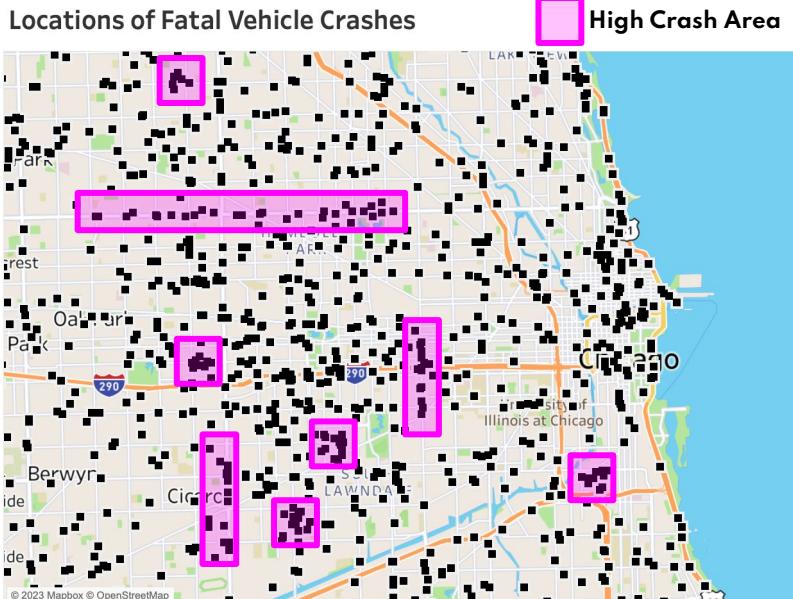
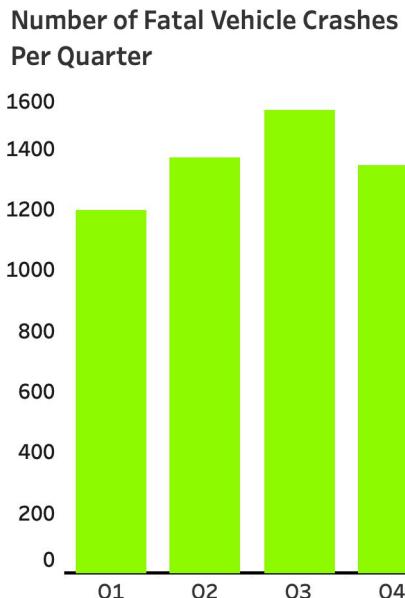
Sex and Age Group



Recommendations:

- Increase availability of fall-prevention programs for seniors
- Increase workplace safety programs and enforcement

Reducing Vehicle Crash-Related Accidental Deaths



Recommendations:

- Road improvements in most dangerous areas
- Increased speed and safety patrols, especially in summer

Binary Classification - Random Forest model

Opioid-Related Deaths

Date/Time /Location Included?	Model Type	Accuracy	ROC
No	Random Forest	82%	0.89

Final Remarks

- Our primary goal:
 - Build a model that could actively guide the Cook County Medical Examiner's Office in better distributing their investigative resources
- Our secondary goal:
 - Use the Office's unique data to provide the City with actionable ideas to reduce accidental deaths





Thank You!

Any Questions?

Sources:

- **Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014**
- **Cook County Medical Examiner's Office,**
<https://www.cookcountyl.gov/agency/medical-examiners-office>
- **Ria Kulshrestha,**
<https://towardsdatascience.com/latent-dirichlet-allocation-lda-9d1cd064ffa2>
- **Jeremy Gorner, Chicago Tribune**
<https://www.chicagotribune.com/news/breaking/ct-covid-illinois-chicago-spike-2021l228-ky7j3c4qx5aqje7n5ufg6estqq-story.html>
- **Slides Go**
- **Susan - beautiful photographs**
- **GA notes & lessons**
- **Wikipedia,**
<https://en.wikipedia.org/wiki/Chicago>