Understanding and Preventing Drug overdoses: An investigative Approach

Team Members: Blake Bleier, Lia Cappellari, Pauline Ranjan

Overview

Our Team conducted a study to analyze the circumstances surrounding drug overdose cases. The goal is to identify key contributing factors and subsequently provide plausible recommendations to reduce the occurrences of this unfortunate event. Understanding how demographics, locations, and the substances involved link up in drug overdoses is crucial to craft effective prevention and intervention plans. By analyzing these occurrences, we hope to provide insights that can help public health officials, policymakers, and healthcare practitioners roll out specific measures to save lives and build healthier communities. Only the key findings from our analysis are found in this report.

Defining our research

Main Research Question

What multifaceted factors, including geographic trends, temporal variations, and specific substance combinations, contribute to drug overdoses, and how can the understanding of these factors inform targeted interventions to reduce the incidence of fatal overdoses and improve public health outcomes?

Areas of Focus

1. Geographic Trends in Drug Overdose Occurrences:

Geographic Analysis: Utilize location data to identify clusters or hotspots of fatal overdoses. Analyze whether certain regions or communities are disproportionately affected.

2. Temporal Variations in Drug Overdose Occurrences:

Longitudinal Study: Conduct a time series analysis to explore changes in drug overdose occurrences over time. Identify trends, fluctuations, or sudden spikes.

Seasonal or Cyclical Patterns: Investigate whether there are seasonal or cyclical patterns in drug overdoses. Are there particular times of the year or specific events that coincide with increased occurrences?

3. Combinations of Drugs:

Drug Combination Analysis: Examine data to identify common combinations of drugs involved in overdoses. Explore which combinations are more lethal and whether certain substances enhance the toxicity of others.

Polydrug Use Trends: Investigate trends in polydrug use, considering how the simultaneous use of multiple substances contributes to overdose incidents.

Correlation with Demographics: Investigate the correlation between demographic groups and the types of substances involved in overdoses. For example, are certain substances more prevalent among specific age groups or genders?

Datasets

Main Dataset:

Accidental_Drug_Related_Deaths_2012-2022.csv (https://github.com/UC-Berkeley-I-School/project2_cappellari_ranjan_bleier/blob/main/Accidental_Drug_Related_Deaths_2012-2022.csv)

Our investigation hinges on a comprehensive dataset, encompassing variables such as age, sex, race, ethnicity, and geographic details of residence, injury, and death. This dataset also provides crucial information on the substances involved, enabling an in-depth analysis of the diverse factors contributing to fatal overdoses. This dataset only features data from Connecticut.

Supplementary Datasets:

Connecticut Median Household Income City Rank (https://acs2019.ctdata.org/#:~:text=Exploring%20Median%20Household%20IncomePer%20Capita,Darker%20blues%20represent%20higher%20values) This dataset only contains data from 2019.

Population by City Dataset (https://portal.ct.gov/DPH/Health-Information-Systems--Reporting/Population/Annual-Town-and-County-Population-for-Connecticut) The dataset from 2019 was used for this analysis

Household income and population data can be merged with the original dataset by city to understand if income or city population has any impact on drug overdose.

Data Cleaning

The data was loaded into a pandas data frame and the variable names were changed into easily understandable, standard names. The data frame consisted of 10654 records and 48 fields.

Several steps were taken to clean the data. Only those of note are mentioned in this section.

Firstly, data validation was conducted to identify and rectify any inconsistencies or errors in the dataset. This included checking for missing values, outliers, or discrepancies in demographic information. Save for missing values, nothing of note needed to be cleaned here.

Secondly, data formats were validated, and illogical categorical values were rectified. Duplicate entries were also verified. The date column had to be formatted into a standard format to foster time-based analysis. Binary categorical variables that served to indicate if a particular drug was involved in the incident had to be changed to contain only ‘Y’ and ‘N’.

For example, the field “ethanol” consisted of 2877 ‘Y’ values and 1 ‘P’ value. The rest of the records had no value for this field as ethanol was not involved in those incidents. After investigating that individual record, we decided to convert ‘P’ to ‘Y’ as it appeared to be a data entry error. Similar actions were taken for the other binary fields.

```
ethanol
Y    2877
P       1
Name: count, dtype: int64
```

→

```
ethanol
Y    2878
Name: count, dtype: int64
```

Figure 1: Initial cleaning of the "ethanol" field

Lastly, missing data values were handled. For instance, for variables like ethanol, the missing values were imputed with 'N' to make it a proper Boolean field. No entire records were dropped due to missing values.

The main dataset did not require extensive data cleaning in this case.

Data Preparation

New columns were created to allow for deeper analysis. One key type of fields that were created pertain to drug combinations. To do this, several steps were taken.

1. The existing binary variables relating to drug presence were grouped into types of drugs. Certain drugs such as fentanyl were not grouped due to their prevalence and need for investigation. The categories chosen were 'heroin', 'cocaine', 'fentanyl', 'ethanol', 'benzodiazepine', 'Opioids', 'Stimulants', 'Other Depressants', and 'Other'.
2. After this, combinations of these categories were created. Some of the original fields that were sorted into these categories were removed.
3. Only the top 10 prevalent combinations were kept in the dataset as the resulting combinations are copious in number and this research project lacks the scope to examine all possible combinations. Please note that combinations can also contain a singular drug or drug category.

	Column	Count
0	fentanyl	1055
1	cocaine_and_fentanyl	834
2	cocaine_and_fentanyl_and_ethanol	527
3	fentanyl_and_ethanol	490
4	heroin_and_fentanyl_and_Opioids	421
5	cocaine	380
6	cocaine_and_fentanyl_and_Opioids	363
7	heroin	347
8	Opioids	326
9	cocaine_and_fentanyl_and_Other Depressants	319

Figure 2: Drug combination variables (Top 10) found within variable names

Several fields that were determined unnecessary to the scope of analysis such as 'description_of_injury' were omitted from the dataset.

As the median household income and population datasets were mostly clean and of fixed structure, it was merged with the main dataset on the "city" field.

Additionally, races and ages were binned into defined categories. Kindly refer to appendix 1 for the data dictionary of the final, analysis - ready dataset.

Whilst investigating possible correlation between drug overdoses, locations, and income levels, several adjustments were made to derive insights.

Firstly, this report opted to consider “number of overdoses per 1000 residents” within a city. This allows for suitable insights to be gathered as different cities of varying median income levels have a different number of residents and should be normalized for an appropriate comparison. The log of the median household income was selected to allow us to better visualize the spread of incomes.

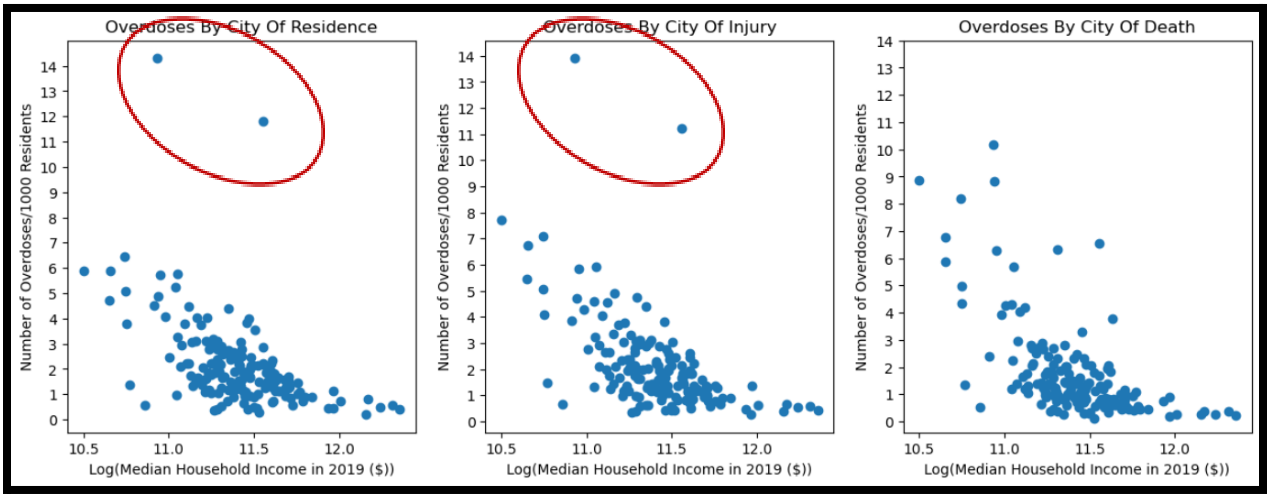

Figure 3: Number of Overdoses/ 1000 Residents by Log (Median Household Income in 2019 ($) for City of residence, injury, and death

Figure 3 displays the correlation between normalized overdoses/1000 residents and the logarithm of median household income across Connecticut cities. Each dot represents a city. All three charts show a noticeable negative linear correlation, indicating that higher household incomes generally correspond to fewer overdoses/1000 residents. The distribution for overdoses by city of residence mirrors that of city of injury, but overdoses by city of death concentrate in a few specific cities. Analysis of death locations reveals a majority of overdoses occurred in hospitals, suggesting victims may have overdosed in or near their town of residence but passed away in specific hospitals.

Additionally, there are two notable anomalous points in both the city of residence and the city of injury plots (highlighted by the red circles in Figure 3), but these points are not as pronounced in the city of death. Further investigation was done into these 2 outliers.

	city	num_overdose	household_income_avg	household_income_std	log_household_inc	est_population	overdose_per_1000
40	EAST HARTFORD	183	55967	3328	10.932518	12800	14.296875
151	WEST HARTFORD	81	104281	3562	11.554844	6869	11.792109

Figure 4: Anomalous points breakdown for residence city

Figure 4 shows the breakdown of the 2 anomalous points in terms of **city of residence**. The cities of interest are East and West Hartford. They have the highest number of overdoses/1000 residents. East Hartford has only half as much household income average as West Hartford.

	city	num_overdose	household_income_avg	household_income_std	log_household_inc	est_population	overdose_per_1000
40	EAST HARTFORD	130	55967	3328	10.932518	12800	10.156250
149	WEST HARTFORD	45	104281	3562	11.554844	6869	6.551172

Figure 5: Stats for cities of interest as city of death

Figure 5 shows the statistics summary for East and West Hartford in terms of **city of death**. Compared against the **city of residence** from Figure 4, the overdose/1000 residents is far lower. It's clear that people from East and West Hartford are overdosing at a high rate, but many of them are not dying while physically located inside East or West Hartford. Interestingly, these cities are geographically positioned to the East and West of Hartford, warranting further investigation into the city of Hartford.

	city	num_overdose	household_income_avg	household_income_std	log_household_inc	est_population	overdose_per_1000
62	HARTFORD	721	36278	1597	10.498967	122105	5.904754

Figure 6: Hartford as City of Residence

	city	num_overdose	household_income_avg	household_income_std	log_household_inc	est_population	overdose_per_1000
62	HARTFORD	1083	36278	1597	10.498967	122105	8.869416

Figure 7: Hartford as City of Death

Figures 6 and 7 reveal that in Hartford 50% more overdose deaths than resident overdoses, indicating a substantial portion of deaths involve non-residents.

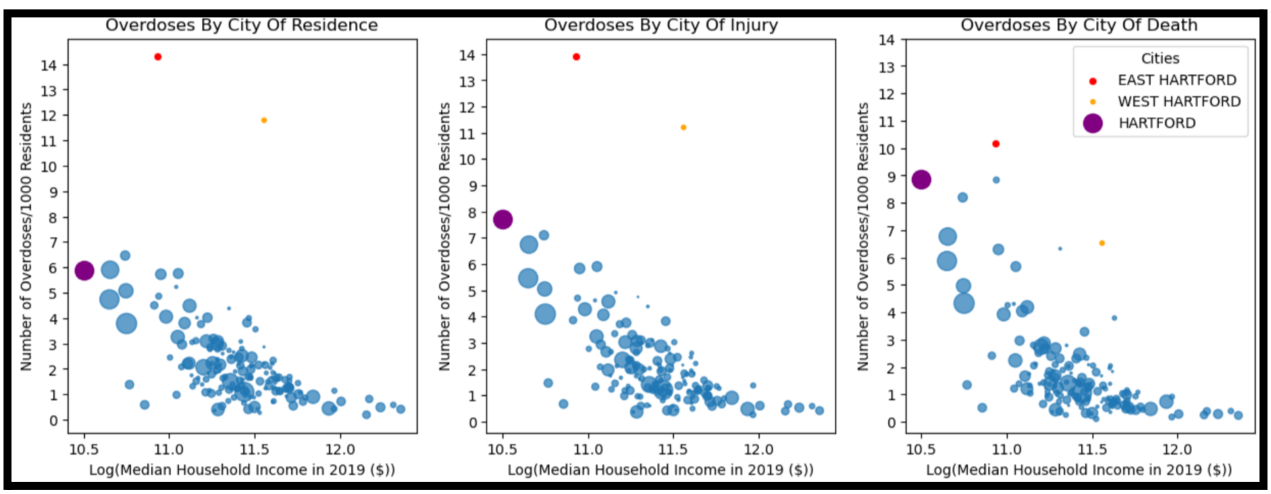

Figure 8: Number of Overdoses/ 1000 Residents by Log (Median Household Income in 2019 ($) for cities of interest

Figure 8 highlights our cities of interest. The size of the bubble represents the total number of overdoses in each city. From this, it is rather clear that Hartford is generally one of the poorest cities and has a high number of overdoses/1000 residents. But it is more pronounced as a city of death in terms of overdoses. This leads to the possible conclusion that residents and people who get injured in East and West Hartford tend to travel to Hartford when they have drug overdoses due to hospital/healthcare facilities that may be more available there.

Though further analysis of the healthcare and transportation setups of the Hartford area is pending, geographically this is a key area of focus in Connecticut.

To investigate temporal variations in drug overdoses, this report considered a timeseries of the total number of overdoses.

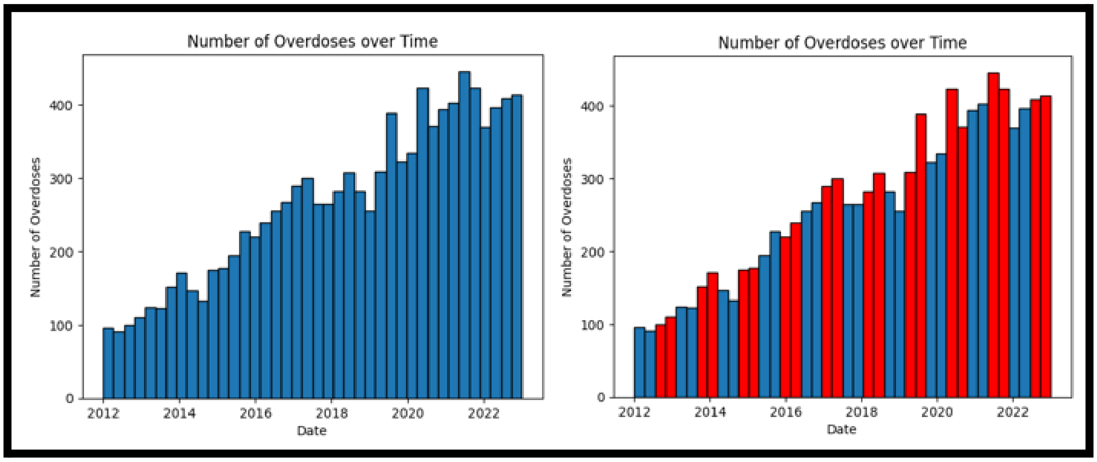

Figure 9: Time Series plot of overdoses over time

As seen in figure 9, the overall trend in sheer number is a steep increase. However, as each bar represents a quarter, we can see that in most years, the later 2 quarters of each year tend to have higher numbers of overdose cases. This report aims to identify a suitable cause for variation observed.

The overdoses pertaining to the identified drug combination variables were overlaid on the time series to check if any of them may have contributed to the observations. Only the relevant charts are shared here.

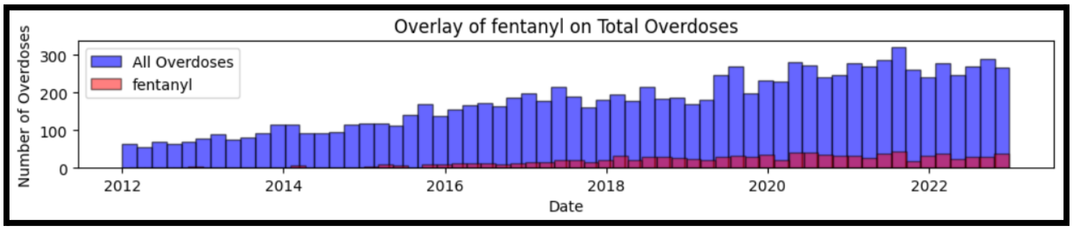

Figure 10: Overlay of overdoses involving only fentanyl over all overdoses

From Figure 10, we can see that an increase in overdoses containing only fentanyl appears to be a major contributor to the increase in overdoses observed from around 2015 to 2022. Some general background knowledge of drug consumption leads us to the conclusion that fentanyl is rather deadly if the user is unaware of the limits. Furthermore, combinations of drugs consumed that contain fentanyl are showing similar trends. Hence, a correlation between larger consumption of fentanyl and overdoses seems like a forgone conclusion.

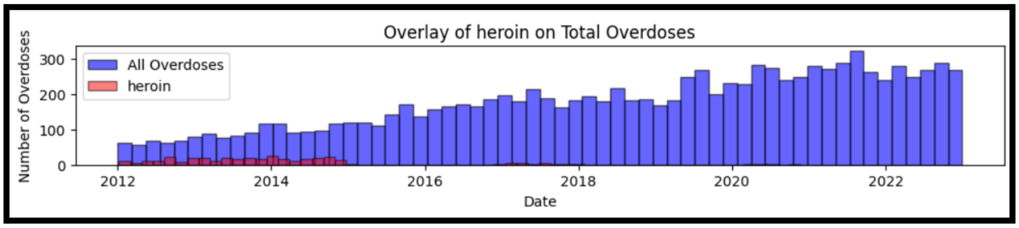

Figure 11: Overlay of overdoses involving only heroine over all overdoses

Figure 11 explains that heroin almost ceased causing overdoses in solidarity at around 2015 which is when fentanyl picked up. Following 2015, there are far more cases of heroin consumed alongside fentanyl and other opioids rather than heroin alone.

Existing research indicates that lacing heroin with fentanyl can make it more lethal. Fentanyl is a potent synthetic opioid, estimated to be 50 to 100 times more potent than morphine and significantly more powerful than heroin. When heroin is adulterated or mixed with fentanyl, it can increase the risk of overdose and contribute to more severe and rapid respiratory depression.

The potency of fentanyl means that even a small amount added to a batch of heroin can have a substantial impact on its effects. Users may not be aware that the heroin they are consuming contains fentanyl, which can lead to unintentional and potentially fatal overdoses.

As for the temporal variations observed in Figure 9, several factors were investigated but nothing within the dataset can truly be noted as the cause for this variation through exploratory data analysis. The next logical step may be to investigate weather data and how it may relate to drug availability, however, it is not within the scope of this report.

Drug Combination Overdose Trends

Further analysis was performed to investigate the top 10 combinations of drugs to identify any key trends pertaining to their prevalence and usage. FIgure 12 shows a ranked order plot of the most common overdose combinations, exclusively. To develop this, overdose numbers for all possible drug combinations (including consumption of singular drugs) were calculated, and the results were rank ordered. The results show that overdoses involving fentanyl dominate the rankings, with fentanyl being the single highest contributor, as well as in 7 of the top 10 causes for overdoses.

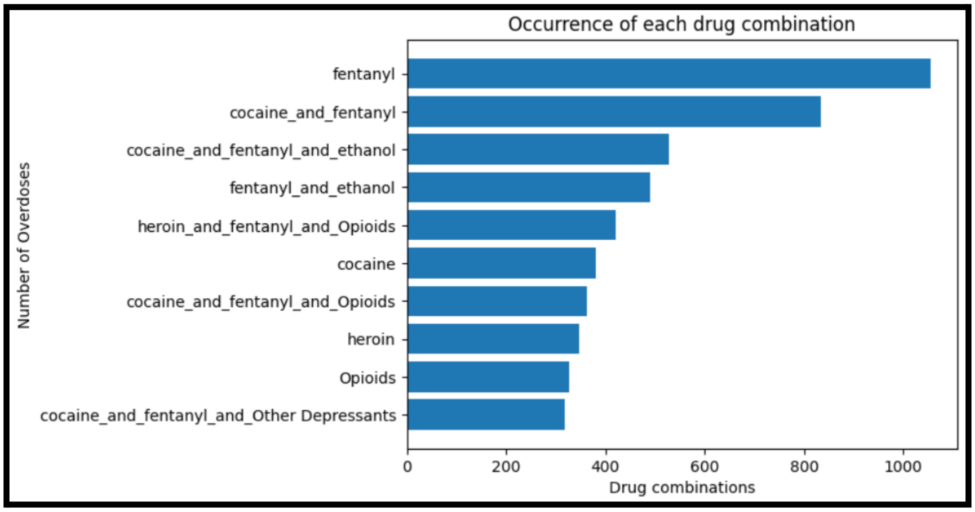

Figure 12: Number of Overdoses of the 10 most prevalent combinations of drugs

To elucidate further, the drug combination overdose data was broken down by race and by age group, as shown in heatmaps in Figure 13 and Figure 14, respectively. For each figure, the values correspond to the probability of overdosing on a combination of drugs given a particular racial demographic or age bracket. Only the top 10 drugs or combinations of drugs were included, and probability rates were normalized to include overdoses only among these top 10 drugs/combinations.

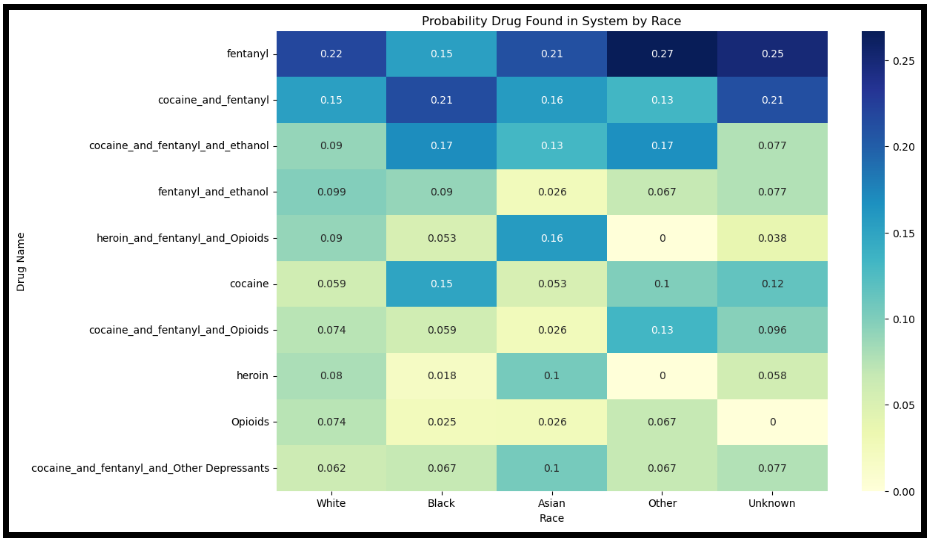

Figure 13: Probability of overdosing on a drug or drug combination given a certain race

Several key observations can be made from Figure 13. Although people of every race tend to frequently overdose on fentanyl by itself, black people have the lowest rates of any racial group. Conversely, they have the highest rates of overdosing on a combination of cocaine and fentanyl, as well as a combination of cocaine, fentanyl, and alcohol. Additionally, they have the highest rates of overdosing on cocaine by itself. It is well documented that in recent years cocaine has been cut with fentanyl by drug dealers, which has led to high numbers of unexpected overdoses due to the highly toxic nature of fentanyl. The data show that the black demographic may be particularly susceptible to this issue relative to other racial demographics. It suggests that those who are black tend to use cocaine more than other races, and are potentially dying most from fentanyl-laced cocaine. We suggest further research into this observation with the intent to ensure proper education and awareness of the dangers of fentanyl.

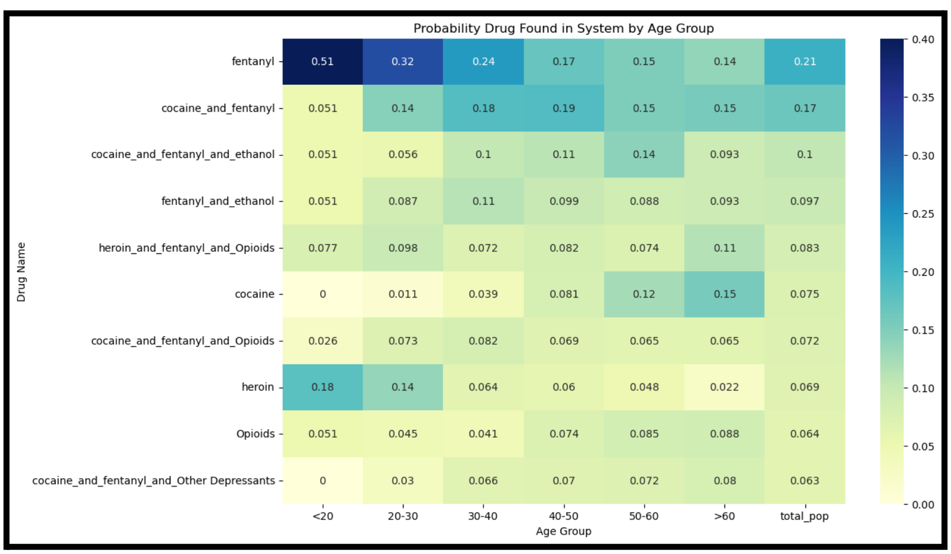

Figure 14: Probability of overdosing on a drug or drug combination given a particular age group

With respect to age demographics in Figure 14, there is a monotonic decrease in the prevalence of overdoses caused by fentanyl alone as age increases. Based on the results, more than half of overdoses in people younger than twenty were caused by fentanyl alone, while only 14% occurred in those older than 60. A similar observation can be made with heroin alone, suggesting that younger demographics are overdosing much more frequently than older generations on fentanyl and heroin. Overdoses on cocaine as a lone drug, however, show a monotonic increase with an increase in age, mirroring the inverse of the fentanyl/heroin results, with only about 1% of overdoses occurring in those aged 30 and below, while 15% of overdoses are due to cocaine for those greater than 60. Those who are middle aged overdose most frequently on a combination of cocaine and fentanyl, again possibly due to a lack of education or awareness of the risk of fentanyl-laced cocaine.

FIgure 15 shows the number of overdoses that occurred in each predefined age bracket. It is worth noting the low count for those younger than 20 years old (39 overdoses). Because of this low sample size, the distribution may be non-representative of the overall population in that age group, however, there are more than the common rule-of -thumb of 30 samples for a large sample set, and can be taken with some degree of confidence. Additionally, there may be a bimodal distribution in frequency with one peak in the 30-40 age group and another in the 50-60 age group, although this could be random noise. We recommend further investigation into this observation on additional datasets to confirm whether this is a repeatable trend or statistical noise.

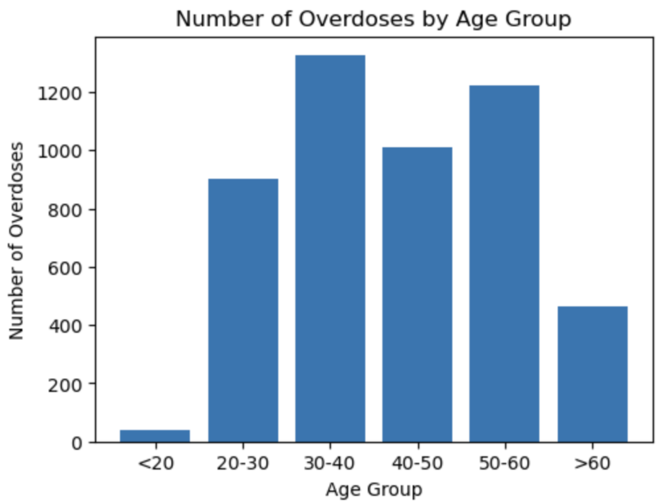

Figure 15: Overdose frequency based on a given age demographic

Consolidated Insights and Recommendations

Insights

1. Geographic patterns in East Hartford, West Hartford, and Hartford reveal fatal outcomes for overdoses, emphasizing the need for targeted interventions.
2. Temporal trends indicate a concerning rise in overdoses linked to the growing preference for fentanyl, notably in combinations with heroin.
3. No clear cause for temporal variation in the third and fourth quarters was identified, requiring further investigation.
4. Top 10 drug combinations highlight fentanyl's prevalence, particularly in seven of the top 10 causes for overdoses.
5. Ethanol appears only in combinations, indicating concern for drug consumption while drinking.
6. Polydrug use significantly contributes to overdoses, with top combinations featuring fentanyl, cocaine, heroin, ethanol, and other opioids.
7. Breakdowns by race reveal lower rates of fentanyl overdoses for black individuals but higher rates for combinations with cocaine, suggesting vulnerability to fentanyl-laced cocaine.
8. Age demographics show decreasing fentanyl overdoses with age, higher rates for younger individuals with fentanyl and heroin, and increasing rates for cocaine in middle-aged individuals.

9. Caution is needed while tailoring programs for those under 20 due to inconclusive data on drug consumption.
10. A potential bimodal distribution in age groups requires further investigation for validation.

Recommendations to public health officials, policymakers, and healthcare practitioners

Geographically Targeted Intervention:

Public health officials should prioritize the development and implementation of targeted intervention programs in East Hartford, West Hartford, and Hartford. Given the observed pattern of individuals traveling from East and West Hartford to Hartford for hospital care following overdoses, initiatives focused on improving accessibility to addiction treatment services and enhancing coordination between healthcare facilities in these regions are paramount.

Address the Growing Preference for Fentanyl:

Recognize and address the concerning increase in overdoses linked to the growing preference for fentanyl. Implement educational campaigns to raise awareness about the risks associated with fentanyl use and promote harm reduction strategies. Consider targeted outreach programs in communities where fentanyl-related overdoses are prevalent.

Investigate Seasonal Variation:

Conduct further investigation into the observed temporal variation in overdoses during the third and fourth quarters of each year. Explore potential factors contributing to this seasonal fluctuation, such as changes in drug supply, socioeconomic conditions, weather conditions, or public events. The findings can inform targeted prevention strategies during specific time periods.

Focus on Polydrug Use Prevention:

Develop comprehensive prevention and education programs addressing the prevalence of polydrug use. Emphasize the risks associated with combinations of substances, especially those involving fentanyl, cocaine, heroin, and ethanol. Tailor awareness campaigns to different demographic groups, considering their specific patterns of drug use.

Addressing Drug Consumption While Drinking:

Recognize and address the heightened risk of combining drugs with alcohol, especially evident in the presence of ethanol in combinations. Implement targeted educational campaigns and collaborate with treatment centers to address this specific pattern of substance use. Community initiatives should raise awareness about the dangers of concurrent drug and alcohol consumption, ensuring a comprehensive approach to substance abuse prevention and intervention.

Age-Specific Prevention Strategies:

Recognize the age-specific patterns in overdoses and tailor prevention strategies accordingly. For younger age groups, focus on education about the risks of fentanyl and heroin use. For middle-aged individuals, particularly those in the 30-40 and 50-60 age groups, emphasize awareness of the dangers of combinations involving cocaine and fentanyl.

Appendix 1: Data Dictionary (Fields used in Analysis)

#	Data Field	Data Field Description
1	date	Date Value
2	date_type	Indicates if date value is date of death or date of death reported
3	age	Age of person upon death
4	sex	Sex of deceased person
5	race	Race of deceased person
6	ethnicity	Ethnicity of deceased person
7	residence_city	City of residence of deceased person
8	injury_city	City where injury occurred
9	injury_place	Place where injury occurred
10	death_city	City where death occurred
11	heroin	Indicator if drug combination was present
12	fentanyl	Indicator if drug combination was present
13	cocaine_and_fentanyl	Indicator if drug combination was present
14	cocaine_and_fentanyl_and_ethanol	Indicator if drug combination was present
15	fentanyl_and_ethanol	Indicator if drug combination was present
16	heroin_and_fentanyl_and_Opioids	Indicator if drug combination was present
17	cocaine	Indicator if drug combination was present
18	cocaine_and_fentanyl_and_Opioids	Indicator if drug combination was present
19	heroin	Indicator if drug combination was present
20	Opioids	Indicator if drug combination was present
21	cocaine_and_fentanyl_and_Other Depressants	Indicator if drug combination was present
22	age_binned	Age group defined in 10 year buckets

*Avg household income was calculated separately for city of residence, injury and death and only used when necessary.