DAS732: Data Visualization Assignment 1 Report Team name: Kurtosis

Sourav M Dileep
IMT2022033
sourav.dileep@iiitb.ac.in

Soham Pawar
IMT2022127
soham.pawar127@iiitb.ac.in

Siddharth Ayathu
IMT2022517
siddharth.ayathu@iiitb.ac.in

I. Dataset

In this assignment, we study the Kaggle Drug Overdose Deaths dataset. The dataset originally describes 5105 cases of drug overdose deaths which took place in Connecticut between years 2012 and 2018. This includes the location, sex, race and other important factors for the cases. We try to find the trends in the data, among the different data fields using various visualizations. The fields present in the dataset are:

1) **ID**: This column's values are used to uniquely identify the cases (represented by rows).
2) **Date**: Date when the incident took place. The type of incident being described by column 'DateType'.
3) **DateType**: Type of incident; 1 for Date Reported and 0 for Date of Death.
4) **Age**: A float value representing the age of the subject.
5) **Sex**: String value representing the sex of the subject.
6) **Race**: String value representing the race of the subject.
7) **ResidenceCity**: String value representing the residence city of the subject.
8) **ResidenceCounty**: String value representing the residence county of the subject.
9) **ResidenceState**: String value representing the residence state of the subject.
10) **DeathCity**: String value representing the death city of the subject.
11) **DeathCounty**: String value representing the death county of the subject.
12) **Location**: The location of the reported incident. These are general locations .eg. Hospital, Residence, etc.
13) **LocationifOther**: A more detailed location value for the items with value 'Other' in column 'Location'.
14) **DescriptionofInjury**: Brief description of the cause of injury .eg. inhalation, substance abuse, etc.
15) **InjuryPlace**: The type of location where the subject was injured .eg. residence, hotel, etc.
16) **InjuryCity**: String value representing the city where the subject was injured.
17) **InjuryCounty**: String value representing the county where the subject was injured.
18) **InjuryState**: String value representing the state where the subject was injured.
19) **COD (Cause of Death)**: The values specify a detailed description of the cause of death of the subject.
20) **OtherSignificantFactors**: This field mostly has null values. The non-null values add to the description of Cause of Death values for the subject(s).
21) **Heroin**: Binary value representing if Heroin was involved in the incident; 1 representing if the drug was present, 0 if it wasn't.
22) **Cocaine**: Binary value representing if Cocaine was involved in the incident; 1 representing if the drug was present, 0 if it wasn't.
23) **Fentanyl**: Value representing if Fentanyl was involved in the incident; 1 representing if the drug was present, 0 if it wasn't. Some values are about the form in which the drug was consumed.
24) **Fentanyl_Analogue**: Binary value representing if any Fentanyl analogue was involved in the incident; 1 representing if the drug was present, 0 if it wasn't.
25) **Oxycodone**: Binary value representing if Oxycodone was involved in the incident; 1 representing if the drug was present, 0 if it wasn't.
26) **Oxymorphone**: Binary value representing if Oxymorphone was involved in the incident; 1 representing if the drug was present, 0 if it wasn't.
27) **Ethanol**: Binary value representing if Ethanol was involved in the incident; 1 representing if the drug was present, 0 if it wasn't.
28) **Hydrocodone**: Binary value representing if Hydrocodone was involved in the incident; 1 representing if the drug was present, 0 if it wasn't.
29) **Benzodiazepine**: Binary value representing if Benzodiazepine was involved in the incident; 1 representing if the drug was present, 0 if it wasn't.
30) **Methadone**: Binary value representing if Methadone was involved in the incident; 1 representing if the drug was present, 0 if it wasn't.
31) **Amphet**: Binary value representing if Amphet was involved in the incident; 1 representing if the drug was present, 0 if it wasn't.
32) **Tramad**: Binary value representing if Tramad was involved in the incident; 1 representing if the drug was present, 0 if it wasn't.
33) **Morhpine_NotHeroin**: Binary value representing if Morphine was involved in the incident; 1 representing if the drug was present, 0 if it wasn't.
34) **Hydromorphone**: Binary value representing if Hydro-

morphone was involved in the incident; 1 representing if the drug was present, 0 if it wasn't.

35) **Other**: This column has text values which are names of drugs used other than the ones which have separate columns for them.
36) **OpiateNOS**: Binary value representing if an Opiod, not otherwise specified, was involved in the incident; 1 representing if the drug was present, 0 if it wasn't.
37) **AnyOpioid**: Binary value representing if any Opiod was involved in the incident; 1 representing if the drug was present, 0 if it wasn't.
38) **MannerofDeath**: This column's value represents the way in which the death happened .eg. if the death was accidental or natural, etc.
39) **DeathCityGeo**: The values of this column contain the Latitude and Longitude values of the location of death of the subject.
40) **ResidenceCityGeo**: The values of this column contain the Latitude and Longitude values of the location of residence of the subject.
41) **InjuryCityGeo**: The values of this column contain the Latitude and Longitude values of the location of injury of the subject.

II. Data Pre-Processing

1) The columns 'DeathCityGeo', 'ResidenceCityGeo' and 'InjuryCityGeo' contain the coordinates of the incident. Every cell contains the coordinates along with the location names or codes. This is present in string format and is formatted into two columns; for latitude and longitude, thus six new columns in total.
2) The COD (Cause of Death) column contains string-based description of the cause of death. This column required some language-based processing for extracting important information for the drugs involved. This was done by first removing 'stop' words from the sentences and then performing `NER (Named Entity Recognition)` on the sentences. The words of interest had the tags 'NN', 'JJ' and 'NNS'. The 'NN' and 'NNS' tags are used to identify Noun words and 'JJ' for Adjective words. This was done using the `spacy` library.
   It was expected that the words with 'NN' and 'NNS' tags would be the drug names. However, these still contained some un-related words concatenated with the drug names. For filtering the data further, the `scispacy` library was used for identifying drug names as this library is used for bio-medical words processing. After this, a list of words was formed with only drug names, across all the rows in the original dataset.
   After this, these drugs were categorized into six categories: Amphetamines, Benzodiazepines, Opiates, Barbiturates, Antidepressants and Antihistamines. These are used in different visualizations later.
3) Data pre-processing involved handling missing values based on specific thresholds. Columns with 40% or more null values were removed entirely to eliminate features with excessive missing data, ensuring cleaner and more reliable insights. For columns with 1% or fewer null values, we opted to remove the corresponding rows to retain the feature while minimizing the loss of data. This approach balanced the need to preserve important features while maintaining data quality for analysis.
4) Removed columns 'DateType' and 'ID' which were not related to the context of the data. The column 'DateType' represents the type of incident when it was reported on that date; if it is the date it was reported or if it is the date of death of the subject. The column 'ID' contains values for unique representation of the cases in the dataset.
5) For the columns 'DeathCounty', 'DeathCity', 'ResidenceCounty' and 'ResidenceCity', data imputation was done in case of missing values. This was done based on the frequency of values across other cells for that column. For instance, if for the value 'STRATFORD' in 'DeathCity', the value 'FAIRFIELD' has occurred in 'DeathCounty' for seven times for other cells and a new cell is now encountered in this column, then the value 'FAIRFIELD'is used.
6) To create the area plot of drug overdose cases by year for each drug, we started with dates formatted as "dd/mm/yyyy." Since we only needed the year, we extracted it by taking the year part of the date string. Then, for each record in the dataset, we checked if the year fell within the range of 2012 to 2018. If it did, we incremented the count of overdose cases for that particular drug in that specific year by 1. This way, we accumulated the total number of overdose cases for each drug over the years. Finally, the organized data allowed us to generate an area plot to visually represent how the overdose cases for different drugs evolved year by year.

III. Objectives

The objective of this analysis is to gain deeper insights into the deaths and injuries of the people because of drug overdose and how it is related to other factors. Specifically, we aim to understand the correlation with the following factors:

1) Deaths, Year and Age based analysis
2) Deaths, Location and types of Drugs based analysis
3) Deaths, Race and Sex based analysis

IV. Data Stories

A. Deaths, Year and Age based analysis

Hypothesis 1: Younger people are more correlated to drug overdose deaths as compared to older people. The idea behind this hypothesis is that younger individuals tend to consume more drugs, leading to higher overdose death counts. Since younger people, defined as those under 30 years of age, are often more involved in risky behaviors, we expect them to have a greater share of drug overdose deaths.

Research also suggests that while younger people may use substances like cocaine or ethanol more often, these drugs are often less lethal compared to harder drugs like opioids. This could explain why overdose death rates for younger individuals may sometimes be lower, despite high rates of consumption [1].

This hypothesis is verified visually, as scatter plots indicate that younger people have fewer deaths compared to older age groups, and drugs like cocaine and ethanol have led to fewer deaths overall.

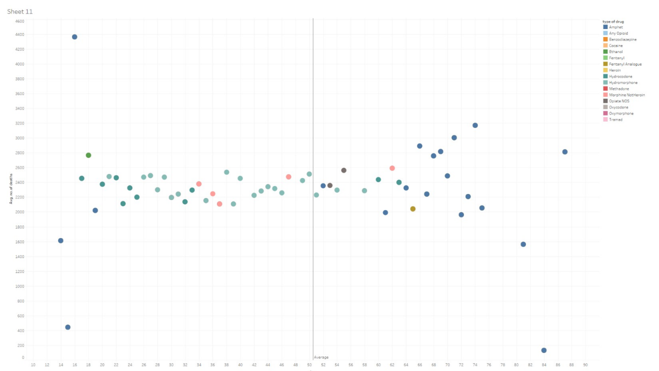

Fig. 1: Scatter plot of average number of deaths by age, with drug categories color-coded and an age average line

This scatter plot displays the average number of deaths versus age, with drug categories represented by color-coded dots. An average age line is included for reference. The plot highlights that Amphetamines (Amphet) were the most lethal, contributing to the highest number of deaths compared to other drugs. This visualization allows for the assessment of drug-related mortality across different ages, with Amphetamines showing a particularly significant impact.

We can verify that younger people tend to take less lethal substances like alcohol, cigarettes, and marijuana, according to sources such as Addiction Center (https://www.addictioncenter.com/addiction/young-adults/). This aligns with the dataset, where the plot demonstrates that individuals below the age of 25 have died from the intake of fewer drugs, supporting the conclusion that younger populations tend to engage with less harmful substances.

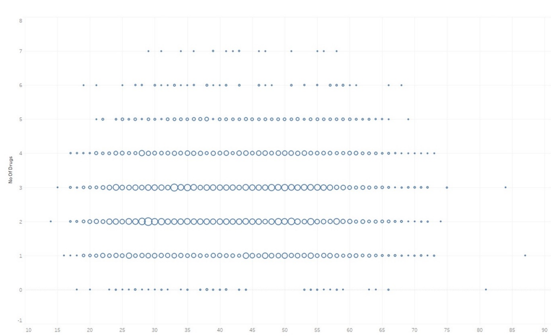

Fig. 2: Circle plot showing the age of drug users and the number of drugs used

This circle plot visualizes the relationship between the age of drug users and the number of drugs they use. Each circle's position reflects a specific age and the number of drugs consumed, while the radius of each circle represents the

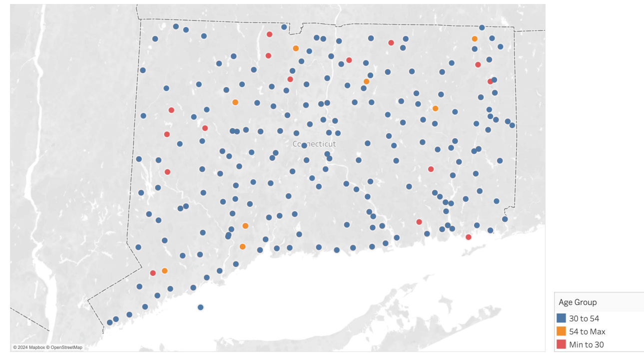

Fig. 3: Geographical Distribution of Drug Overdose Deaths by Age Group in Connecticut

number of users for that age and drug combination. Larger circles indicate higher numbers of users, helping to identify the most common age groups and their corresponding drug usage patterns.

Another reason supporting this hypothesis is that younger people generally have healthier bodies compared to older individuals, which may contribute to fewer deaths despite drug use. Younger bodies may have a greater ability to metabolize and recover from the substances consumed, leading to a lower fatality rate from drug misuse.

This cartograph represents the geographic distribution of drug overdose deaths across Connecticut, segmented by age groups. Three age categories are used: individuals aged under 30 (in red), aged 30 to 54 (in blue), and aged 54 or above (in orange). Each point marks the location of a recorded death, with color coding indicating the corresponding age group. The map provides insight into how drug-related deaths are distributed across the state, with the majority of cases falling in the 30 to 54 age group.

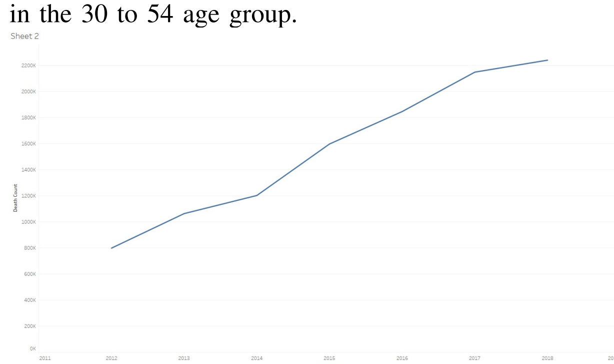

Fig. 4: Line plot showing the irregular increase in deaths from 2012 to 2018.

The data reveals that the number of deaths has increased irregularly from 2012 to 2018. This trend is consistent with the dataset, which shows fluctuations in death rates over this period. The irregular pattern indicates variability in the number of deaths year by year, aligning with the observed data trends. Now, let's explore how the number of drug overdose deaths has changed for each specific drug over the years, building on the overall trend we've just seen. This will help us better understand how each drug contributes to the total number of deaths.

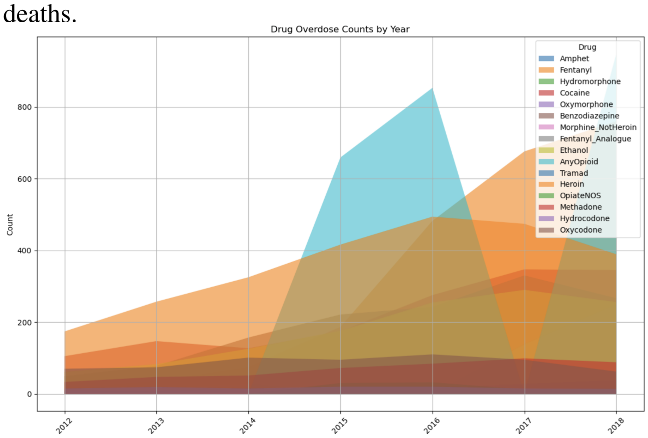

Fig. 5: Drug overdose counts by Year for various substances (2012–2018)

This area plot visualizes the increasing number of drug overdose deaths over the years 2012 to 2018, categorized by the type of drug involved. The x-axis represents the years, while the y-axis shows the total number of overdose counts for each drug, represented by different colors. Notable trends include the sharp rise in fentanyl-related overdoses (shown in blue) around 2015, which continues to increase through 2017. Other opioids like heroin, oxycodone, and methadone are also represented, although fentanyl clearly dominates the latter part of the timeline. The chart suggests a significant shift towards more potent opioids driving overdose death counts in recent years.

B. Deaths, Location and types of Drugs based analysis

There are 6 columns in the dataset that provide insights into residence, death location, and injury place. These columns are ResidenceCity, ResidenceCounty, DeathCity, DeathCounty, Location (of death), and InjuryPlace. All the locations in the filtered dataset fall within 8 counties, all of which are located in the state of Connecticut (CT), USA. from figure 5, it is clear that most of the injuries have taken place in residences, but low average amount of drug taken per person who died.

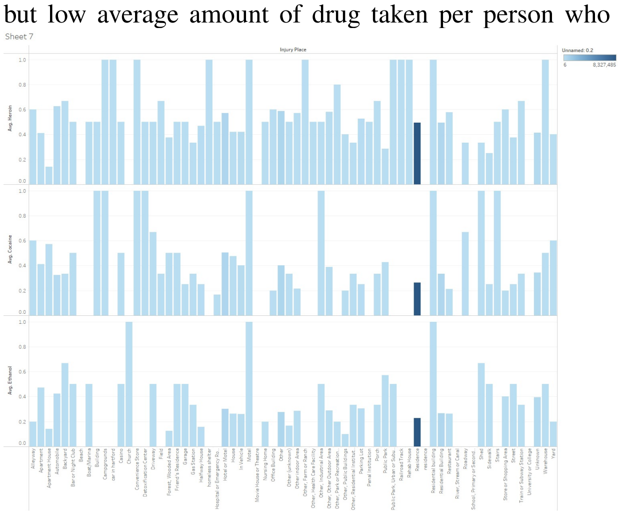

Fig. 6: Bar graph of drug-related deaths by location and drug type

This bar graph compares the number of deaths from the three most widely used drugs across various locations. The depth of the blue color indicates the death count, with a deeper blue representing higher counts. Residences show the highest number of deaths, significantly outnumbering other locations, though the average level of drug abuse per person is lower in these areas compared to others. This visualization highlights the correlation between drug usage and death locations, emphasizing the prominence of residential deaths. From Figure 5, we observe that the majority of injuries due to drug abuse have occurred in residences, despite the fact that the average quantity of drugs consumed per injured person in this setting is lower than anticipated and also lower compared to other venues. The following plot highlights that a greater variety of drugs has been consumed in residences, which reduces the average quantity of any single drug taken at home. This suggests that the presence of multiple drug types might contribute to the pattern of injury, even though individual drug consumption levels remain modest.

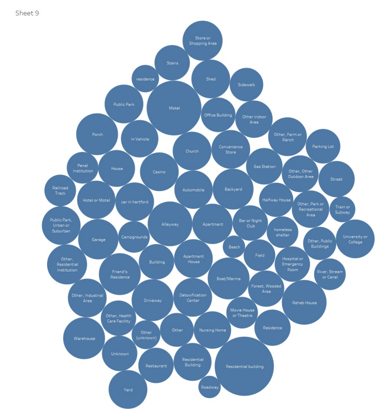

Fig. 7: Circle plot of injury locations with circle size representing different drug types used.

The circle plot displays the locations where injuries have occurred. In this plot, the size of each circle corresponds to the number of different types of drugs used at each location, rather than the total quantity of drugs used. This visualization helps to understand the diversity of drug types involved in injuries across various locations.

Hypothesis 2: It is expected that individuals from wealthier areas are more prone to engaging in substance abuse, and consuming more drugs results in a higher likelihood of death due to overdose.

Research [5] has indicated that affluent communities tend to experience elevated rates of substance abuse, particularly among adolescents and young adults. This is largely attributed to factors such as heightened academic pressure in elite schools, where students are often held to high expectations of achievement. Stress and competition in these environments can lead to the misuse of substances like ethanol, marijuana, and stimulants (e.g., Adderall and cocaine), which are more accessible to wealthy teens due to their greater disposable income.Additionally, easier access to drugs and alcohol, along with social influences where substance use may be normalized within certain peer groups, contributes to the trend. Wealthier teens may engage in drug use as part of a cultural or social dynamic within their communities, which can lead to higher rates of drug abuse and overdose fatalities.

Hartford and New Haven contain areas of both wealth and poverty, reflecting the broader economic disparities seen in Connecticut. Hartford is often associated with affluence due to high-earning neighborhoods and the presence of wealth. Greater Hartford ranks high in terms of the percentage of top-earning households, with 26% of these households living in predominantly affluent. Similarly, in New Haven [6], wealth is concentrated in certain neighborhoods. The top earners in the New Haven-Milford area have an average income of $282,000, placing it among the top 20 regions in the country for high-income households.

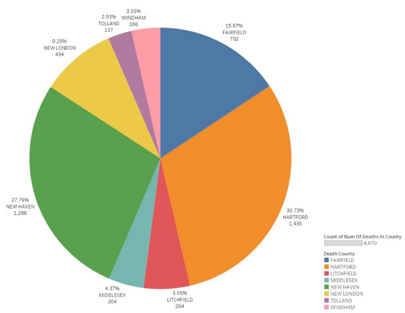

Fig. 8: Pie chart of death distribution by county.

This pie chart represents the distribution of death counts by county in Connecticut. The largest segment is Hartford County, accounting for 30.73% of the deaths, followed closely by New Haven County with 27.79%. Fairfield County holds 15.67%. Other counties like New London, Middlesex, Litchfield, Windham, and Tolland make up the remaining portions with smaller percentages. This chart visually emphasizes that the majority of deaths are concentrated in Hartford, New Haven, and Fairfield counties.

Thus we can see from fig 7 that Hartford, New Haven and Fairfield , the three richest counties in Connecticut have the most drug overdose deaths.

From the below tree map we can also see that the same counties have highest no.of drugs consumed as well.

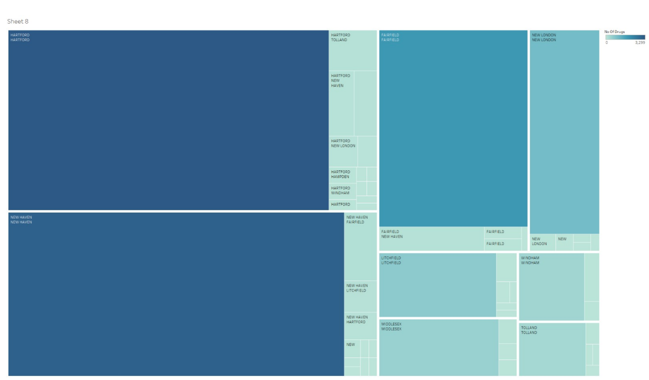

Fig. 9: Treemap of deaths by county and city, with cell size representing death count and shade indicating number of drugs used.

The treemap visualizes the number of deaths by county and city, with each cell representing a specific location. The size of each cell corresponds to the number of deaths in that area, while the color shade indicates the number of different drugs used. Darker shades of color are associated with larger cells, suggesting that higher drug use correlates with a greater number of deaths. This visualization highlights the relationship between drug use and mortality across various locations.

The analysis of deaths caused by drug overdose reveals a striking trend in the types of substances contributing to fatal incidents. The accompanying pie chart visually represents the distribution of drug-related deaths over the years, segmented by the specific drugs involved.

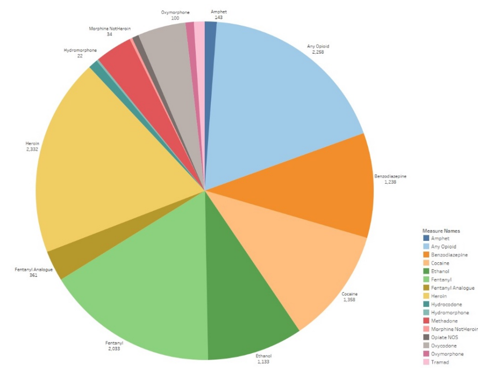

Fig. 10: Pie chart showing the distribution of deaths by drug type.

It is clear from the data that Heroin and Opioids are the predominant contributors to overdose deaths, significantly outpacing other drugs in the dataset. This indicates the severe impact of the opioid crisis, where synthetic opioids like fentanyl, as well as natural and semi-synthetic opioids such as heroin, have become the primary cause of fatalities. This chart underscores the critical public health challenge posed by these drugs, illustrating that efforts to combat overdose deaths must focus heavily on opioids. The visualization also highlights the disproportionate rate of deaths caused by heroin and opioids when compared to substances like cocaine, ethanol, or marijuana. Despite the widespread use of these other drugs, their contribution to the overall death toll is comparatively smaller. By illustrating the significant share of overdose deaths attributed to these substances, this data provides essential insights into the areas where public health interventions and policies need to be directed, particularly in addressing the opioid epidemic. This calls for continued education, prevention strategies, and access to treatment, specifically targeting opioid and heroin abuse, to reduce fatalities and mitigate the ongoing crisis.

One way to categorize drugs is based on the purpose of intake or their intended effects:

The histogram displays the counts of various substances

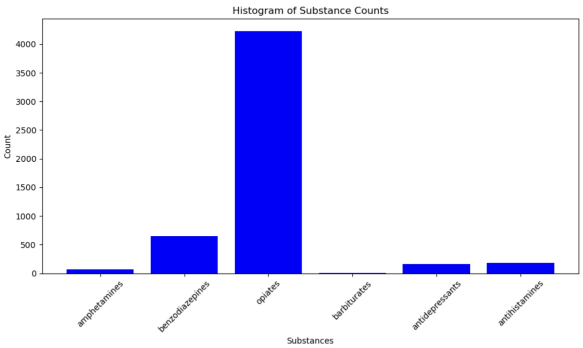

Fig. 11: Histogram of substance involvement in drug overdose cases

involved in drug overdose cases, highlighting their frequency of occurrence. The x-axis lists different substance categories, including amphetamines, benzodiazepines, opiates, barbiturates, antidepressants, and antihistamines, while the y-axis represents the count of cases in which these substances were present. The overwhelming majority of cases involve opiates, as seen by the tall bar dominating the graph, followed by benzodiazepines. Other substances, such as amphetamines, antidepressants, antihistamines, and barbiturates, are far less common in the data.

We can also categorize drugs based on their means of consumption. There are 4 major modes of drug intake by a human, these are:
1.) ORAL - This involves taking drugs through the mouth in the form of pills, liquids, or capsules.

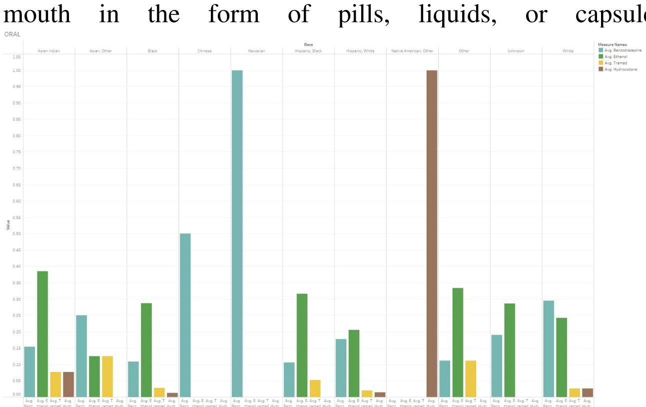

Fig. 12: Bar plot of average oral drug consumption by race.

The bar plot displays the average consumption of four types of oral drugs—Benzodiazepine, Ethanol, Tramad, and Hydrocodone—per person across different racial groups. The data indicates that all Native Americans who died due to drug misuse had used Hydrocodone, while all Hawaiians who died had used Benzodiazepine. Asian Indian shows a notable average for the consumption of Hydrocodone (green bar), with lower averages for Tramad (yellow bar) and Ethanol (orange bar)and Benzodiazepine (blue bar) usage is not significant.
Among Asian,Other Hydrocodone is also the most consumed drug, though the average values are relatively low compared to other races and Ethano, Tramad have lower yet comparable levels.
For Blacks Hydrocodone is again the most prominent drug .Tramad and Benzodiazepine consumption are also observed at moderate levels, with Ethanol having minimal representation.
For the Chinese, the average value for Benzodiazepine consumption is exceptionally high in this group compared to other drugs. This indicates a significant prevalence of this drug within the Chinese group. Other drugs, such as Hydrocodone, Ethanol, and Tramad, are either non-existent or very low in comparison.
For Hawaiians Hydrocodone (brown bar) average consumption is significantly high for this group, representing a dominant drug in this population, they do not have significant consumption of other drugs.
For Hispanic, White Hydrocodone has the highest average consumption, followed by Benzodiazepine. Tramad and Ethanol have moderate and lower consumption rates, respectively.
In the group of Native American, Hydrocodone consumption is overwhelmingly dominant, with no significant averages in other drugs.
The most consumed drug amongst Whites is Hydrocodone, followed by Benzodiazepine and Tramad. Ethanol consumption is present but at a lower average value compared to other drugs.

2.)Injection - Drugs administered via needle, either intravenously, intramuscularly, or subcutaneously.

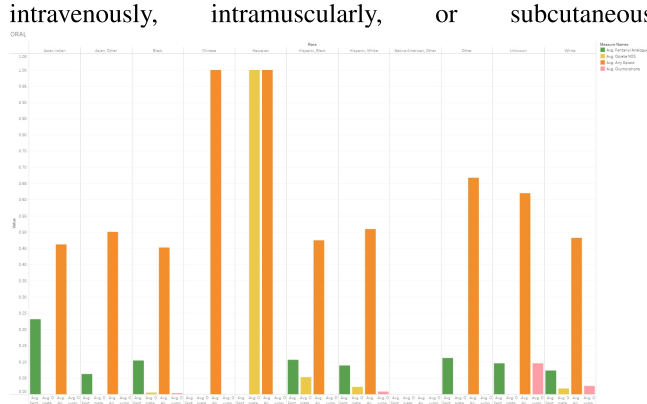

Fig. 13: Bar plot of average injection drug use by race.

The bar chart you provided shows the average oral drug consumption across different racial groups, with specific drugs like fentanyl analogues, oxycodone NOS, and oxymorphone highlighted.
Chinese and Hawaiian populations exhibit the highest average consumption of oral drugs, particularly oxycodone NOS.
Native American and Black individuals also show significant use of oral drugs, but at lower levels compared to the Chinese and Hawaiian populations.
The orange bars, representing oxycodone NOS, are predominant across most racial categories, indicating it is the most frequently consumed oral drug in this dataset.

There are smaller, but noticeable, consumption levels of any opioid and fentanyl analogues, represented by green and yellow bars respectively.

3.)Intranasal (Nasal) - Drugs snorted through the nose, allowing absorption through the nasal mucosa, which can produce a rapid onset of effects as the drug enters the bloodstream quickly.

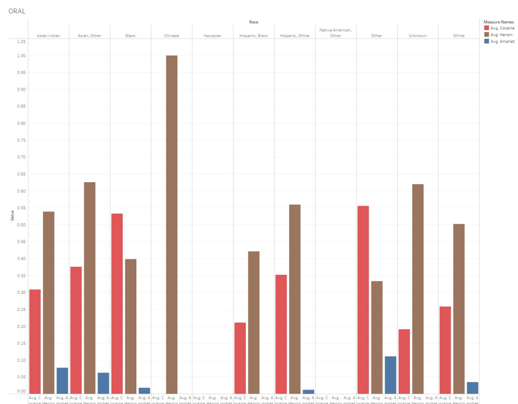

Fig. 14: Bar plot of intranasal drug use by race.

For Asian Indian and Others Heroin is the predominant drug consumed, with notable amounts of cocaine. Amphet consumption is minimal or non-existent in both subgroups.
For Blacks both heroin and cocaine are used in significant amounts. Amphet usage remains quite low, almost negligible compared to the other drugs.
For the Chinese, Heroin consumption is very high, making it the most consumed drug in this group. Cocaine and amphet usage are minimal or absent in this group.
For Hispanic Blacks and Hispanic Whites, Heroin is consumed in the highest quantities, followed by moderate levels of cocaine. Amphet is present, but its usage is quite low across both groups.
For Whites the most significant consumption is of heroin, followed by a considerable amount of cocaine. Amphet usage is lower but still present compared to other racial groups.

4.)Buccal: Drugs placed between the gum and cheek (buccal area) or under the tongue (sublingual) to be absorbed directly into the bloodstream through the oral mucosa.

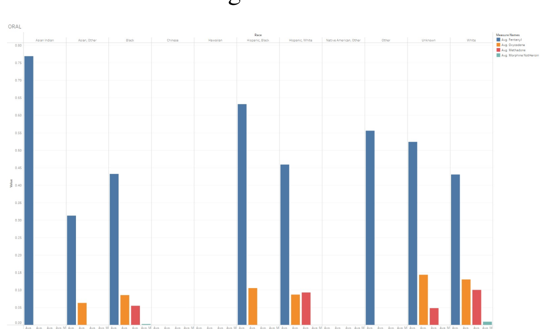

Fig. 15: Bar plot of buccal drug use by race.

Asian Indian: Fentanyl appears to be the predominant drug used, with its usage much higher than other substances. Oxycodone and Methadone have similar, though much lower, values compared to Fentanyl.
Asian, Other: There is a notable consumption of Fentanyl, which dominates usage compared to the other drugs in this group. However, Oxycodone and Methadone have some, but lower, reported use.
Black: Fentanyl is once again the most consumed drug, though the values are lower compared to the previous groups. Methadone and Oxycodone are consumed in smaller, but still noticeable amounts.
Hispanic, Black: The consumption of Fentanyl remains significant, with much smaller amounts of Oxycodone and Methadone reported. Morphine usage in this group is minimal.
White: Fentanyl leads the chart again, being the most consumed drug. The consumption of Oxycodone and Methadone is slightly higher compared to the other racial groups, but still significantly less than Fentanyl. Morphine consumption is minor.
Therefore, Fentanyl stands out as the most prevalent drug in terms of average use across all racial groups. Oxycodone and Methadone also show some usage but in smaller quantities. Morphine has the least amount of usage across the dataset, regardless of the racial group.

C. Deaths, Race and Sex based analysis

The columns sex,race are very important to learn realistic insights about the deaths due to drug overdose.
Hypothesis 3: It is a common misconception that Black people use more drugs than other racial groups. In fact, drug use rates among Black and white Americans are very similar. However, Black Americans are significantly more likely to be arrested for drug-related offenses—about 2.7 times more likely than white Americans, despite similar usage rates [7].
But from the below figure, we can see a contradiction to this stereotype. The data actually suggests that more deaths due to drug overdoses occur among white individuals than Black individuals. This discrepancy could stem from several factors. One possibility is that the opioid crisis, which has disproportionately impacted white, suburban, and rural communities, plays a significant role in these overdose statistics. Furthermore, socioeconomic factors and disparities in healthcare access may lead to higher overdose mortality among white populations. This stands in contrast to public perception, illustrating that drug-related issues affect all racial groups and that overdose deaths, particularly in recent years, have been more prevalent in white populations due to the widespread availability and misuse of prescription opioids, fentanyl, and other synthetic drugs.
Additionally, while Black Americans face higher rates of incarceration for drug offenses, they are not necessarily engaging in higher levels of drug use or experiencing higher levels of overdose deaths compared to other groups. This challenges the traditional stereotype and highlights the need

to address drug abuse and overdose as a broader societal issue, rather than one confined to any single racial group.

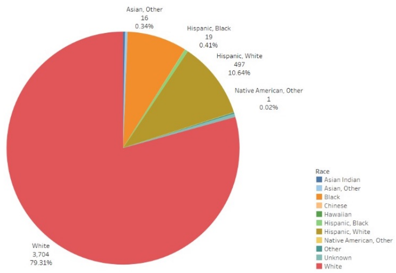

Fig. 16: Pie chart showing the distribution of drug-related injuries by race.

The pie chart illustrates the distribution of drug-related injuries across different racial groups. It reveals that White individuals constitute the majority of these injuries, accounting for 79% of the total. In contrast, racial groups such as Chinese, Native Americans, Hawaiians, Asians, Asian Indians, and others each represent less than 1% of the total injuries. This chart effectively highlights the significant disparity in drug-related injuries among different races.

We can reconfirm this observation by analyzing the trends in the below line graph, which further supports our hypothesis. The graph shows a clear distinction, with white individuals accounting for a larger proportion of deaths due to drug overdose compared to Black individuals. This directly contradicts the stereotype that Black people use more drugs or are more heavily involved in substance abuse.

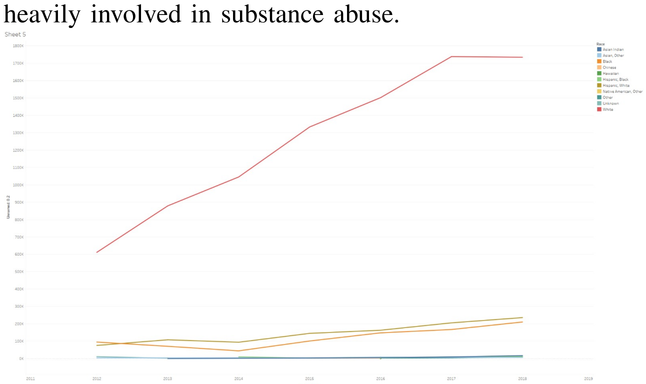

Fig. 17: Line plot showing yearly drug-related deaths by race, highlighting trends and comparisons.

The analysis shows a significantly higher number of drug-related deaths among White individuals compared to other racial groups across all years. The data indicates a sharp increase in deaths among Whites each year. Notably, the number of deaths among Whites remained consistent between 2017 and 2018. Additionally, the death counts for Hispanic and Black individuals are relatively close to each other throughout the years, indicating similar trends for these racial groups.

Therefore, the hypothesis—"It is a common misconception that Black people do more drugs than other racial groups"—has been disproven. The data shows that white people, particularly in the context of the opioid crisis, suffer more from drug overdose fatalities. This underscores the need to reshape public perceptions and address substance abuse as a complex issue affecting various demographics, rather than reinforcing harmful racial stereotypes.

Next, we observe an even distribution of deaths within the white population across the state of Connecticut due to drug overdose. This suggests that, unlike certain racial or ethnic groups where overdose deaths may be more localized or concentrated in specific regions, the issue of substance abuse and its fatal consequences affect the white population more uniformly throughout the state. This even spread indicates that drug misuse is a widespread problem across various communities, irrespective of location, further reinforcing the narrative that substance abuse transcends racial and geographical boundaries.

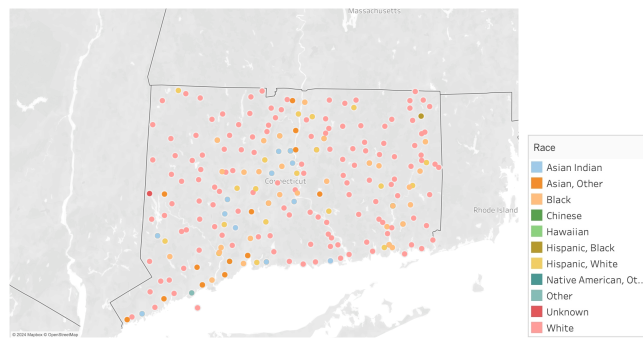

Fig. 18: Cartograph representing the race of death case subjects and their geographic distribution across the state

The cartograph in the report provides a visual representation of the racial breakdown of drug overdose deaths across Connecticut, linking each case to its geographic location. This visualization offers insight into the spatial and demographic distribution of fatalities, highlighting potential disparities among different racial groups. Each race is represented by distinct colors or symbols, allowing for easy identification of patterns across the state. Notably, the map may reveal geographic clusters where certain racial groups experience a higher concentration of overdose deaths, offering clues to regional disparities in health outcomes and access to services. The visualization underscores differences in urban and rural impacts, helping to inform public health efforts aimed at addressing these disparities in overdose fatalities.

Next, we observe the distribution of deaths due to drug overdose between males and females in the state of Connecticut. Data indicates that the number of male overdose deaths is generally higher than that of females, a trend commonly seen in many regions. This disparity can be attributed to several factors, including differences in drug access, societal pressures, and risk-taking behavior that may influence substance use patterns between genders. However, while males show higher fatality rates, it is important to note that female overdose deaths are also significant and have been steadily increasing, emphasizing that drug misuse impacts both genders deeply.

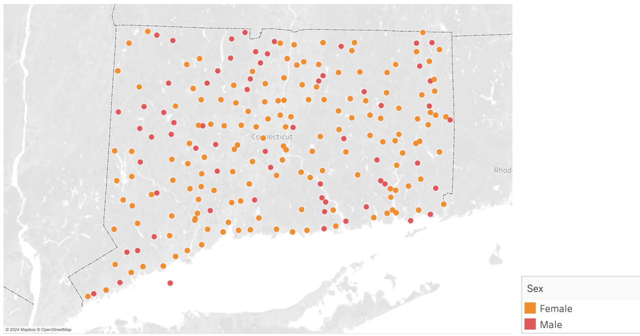

Fig. 19: Drug overdose deaths in Connecticut by sex

The map provides a detailed view of drug overdose deaths across Connecticut, with each point representing a case and color-coded by the sex of the victim—orange for females and red for males. The geographic distribution reveals significant clusters in southern Connecticut along the coast, as well as in central and eastern parts of the state. This visualization highlights potential regional and gender-based disparities, offering insights into the spatial patterns of overdose fatalities. The clear color distinction between male and female cases allows for a quick demographic analysis, helping public health officials identify high-impact areas for targeted interventions.

Finally, we examine how many different kinds of drugs are consumed by each age group, separated by gender.
Gender-wise, males across most age groups tend to consume a greater variety of drugs compared to females. This aligns with the higher incidence of drug-related deaths observed among men. However, in certain age brackets, particularly in younger cohorts, female drug consumption patterns have become more diverse, signaling a shifting trend in substance use.

Fig. 20: Circle plot illustrating drug use and gender distribution in drug-related deaths by age group

The circle plot visualizes the number of drugs used by individuals who died across different age groups. Each circle in the plot corresponds to an age group, and the size of the circle indicates the number of deaths in that age group involving a specific number of drugs. Most deaths occur among individuals who used 2 or 3 drugs. The plot also highlights a gender distinction: blue circles represent males, and orange circles represent females. In many age groups, the blue circles are larger and form the outer ring around smaller, concentric orange circles, indicating that more males died from drug use compared to females. This pattern suggests that male deaths are more prevalent in most age groups where drug use is involved.

V. Member Wise Contributions

After a detailed discussion about the dataset and understanding the context, we came up with the objectives mentioned in Section III. For the visualizations, objectives and hypotheses, the following distribution was followed:

1) Objective 1 (Deaths, Year and Age based analysis) was primarily done by Soham. Siddharth and Sourav suggested a few relations among the visualizations which were appropriately valid and necessary for the hypothesis. Soham also did the pre-processing of data for the COD column, which required concepts from NLP (Natural Language Processing) to extract useful features from the text. He also did the processing for generating geolocation coordinate columns for the cartographs and 'Date' formatting for the Area plot.
2) Objective 2 (Deaths, Location and types of Drugs based analysis) was primarily done by Siddharth. Soham and Sourav suggested a few relations among the visualizations which help relate the visualizations under the given hypotheses. Siddharth also helped plot several bar graphs and pie charts for the visualizations of Objective 2. He also helped with data cleaning, employing important rules for dropping rows and columns which helped in plotting many important visualizations.
3) Objective 3 (Deaths, Race and Sex based analysis) was primarily done by Sourav. Siddharth and Soham suggested a few relations among the visualizations and were appropriately connected under the hypothesis. Sourav also helped in forming many accurate hypotheses used in the report for all the Objectives. The hypotheses suggested and corrected by him made the visualizations more understandable and the inference easy to see.

References

[1] Juan M Dominguez Ph.D., Psychology Today, "Cocaine Increases Risky Behaviors, Depending on Your Age ". Link
[2] Drug Policy Facts, "Young People and Substance Use". Link
[3] Addiction Center, "Young Adults". Link
[4] EUDA: European Union Drugs Agency, "Young people and drugs". Link
[5] Biomedcentral"adolescents with high socioeconomic status more likely to engage in drugs". Link
[6] NewHavenIndependent"New Haven and Hartford are affluent counties". Link
[7] Hamiltonproject"arresting of blacks vs whites for drug abuse". Link