Team
Data Addicts

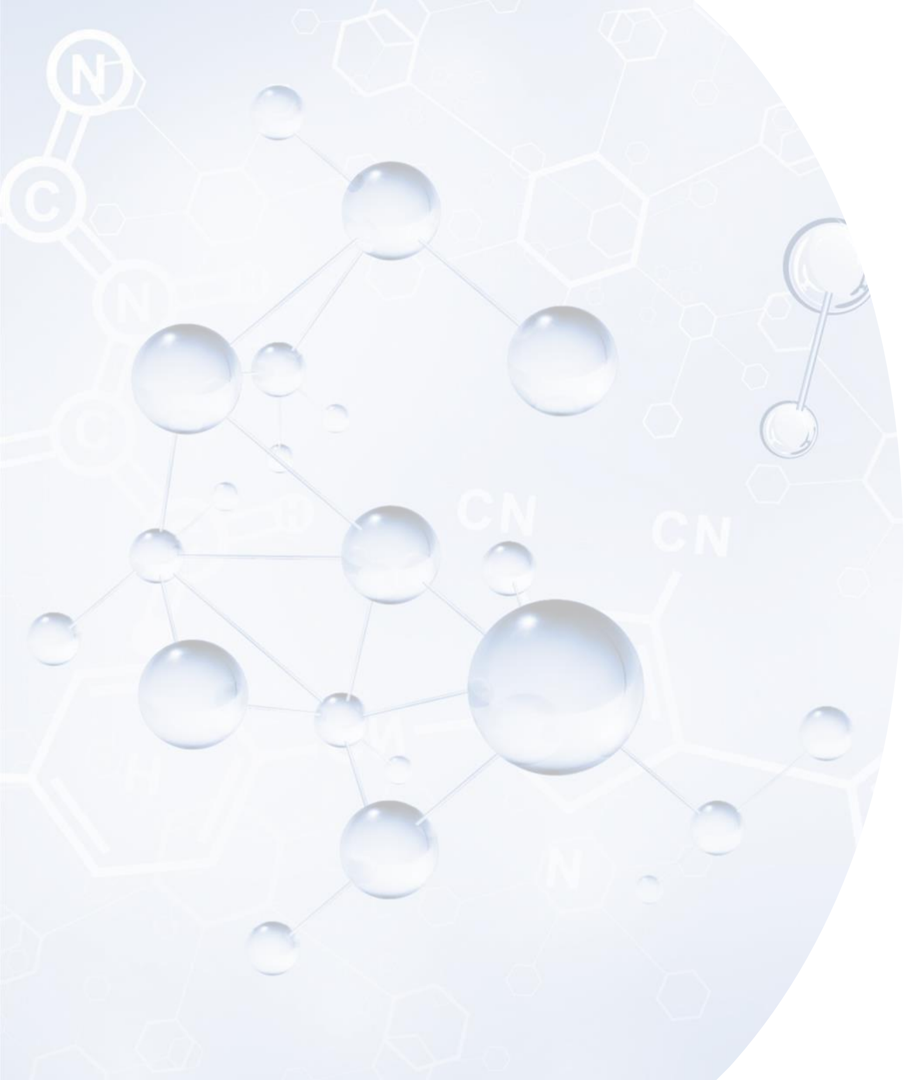
Hannah, Kevin, Steven, Cheryl, Tony

Uncovering the Epidemic: Drug Overdose Deaths in the United States

May 19th, 2024

Outline

Data Preprocessing	What we did to clean and pre-process the data
Data Analysis & Visualization	Insights gained from data analysis
Conclusion	Impacts on health policies



Data Preprocessing

Data Preprocessing

```python
def parse_stub_label(row):
    stub_name = row['STUB_NAME']
    stub_label = row['STUB_LABEL']

    sex, race, hispanic_origin, age = None, None, None, None

    if stub_name == 'Total':
        pass
    elif stub_name == 'Sex':
        sex = stub_label
    elif stub_name == 'Sex and race':
        sex, race = stub_label.split(': ')
    elif stub_name == 'Sex and race and Hispanic origin':
        parts = stub_label.split(': ')
        sex = parts[0]
        race = parts[2]
        hispanic_origin = parts[1] if len(parts) == 3 else 'Non-Hispanic'
    elif stub_name == 'Age':
        age = stub_label
    elif stub_name == 'Sex and age':
        sex, age = stub_label.split(': ')
    elif stub_name == 'Sex and race (single race)':
        sex, race = stub_label.split(': ')
    elif stub_name == 'Sex and race and Hispanic origin (single race)':
        parts = stub_label.split(': ')
        sex = parts[0]
        race = parts[2]
        hispanic_origin = parts[1] if len(parts) == 3 else 'Non-Hispanic'

    return pd.Series([sex, race, hispanic_origin, age])


data[['SEX', 'RACE', 'ETHNICITY', 'AGE']] = data.apply(parse_stub_label, axis=1)
```

To conduct data analysis incorporating all relevant factors, parsing the 'STUB_LABEL' based on the 'Sex', 'Race', 'Hispanic Origin', and 'Age' variables is necessary.

These variables are initially set to *None*. Using the parse_stub_label function, which consists of a series of if-elif statements, new columns ('SEX', 'RACE', 'ETHNICITY', 'AGE') in the DataFrame will be populated with values extracted from the parsed 'STUB_LABEL'.

This function reads the 'STUB_NAME' to determine how to split and assign parts of the 'STUB_LABEL' to the appropriate variables. This approach ensures that demographic information is systematically extracted and organized

Data Preprocessing – Analyzing Trends (All Variables except *Age*) Pt. 1

```
data_new = data.drop(columns=['FLAG', 'AGE'])
data_new.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6228 entries, 0 to 6227
Data columns (total 16 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   INDICATOR       6228 non-null   object
 1   PANEL           6228 non-null   object
 2   PANEL_NUM       6228 non-null   int64
 3   UNIT            6228 non-null   object
 4   UNIT_NUM        6228 non-null   int64
 5   STUB_NAME       6228 non-null   object
 6   STUB_NAME_NUM   6228 non-null   int64
 7   STUB_LABEL      6228 non-null   object
 8   STUB_LABEL_NUM  6228 non-null   float64
 9   YEAR            6228 non-null   int64
 10  YEAR_NUM        6228 non-null   int64
 11  AGE_NUM         6228 non-null   float64
 12  ESTIMATE        5117 non-null   float64
 13  SEX             4908 non-null   object
 14  RACE            2268 non-null   object
 15  ETHNICITY       1272 non-null   object
dtypes: float64(3), int64(5), object(8)
memory usage: 778.6+ KB
```

Dropping the '*FLAG*' and '*AGE*' columns is necessary in this step to analyze the trends of drug overdose deaths over time based on other variables (Sex, Drug Types, Ethnicity, and Race).

The '*FLAG*' column identifies the presence of null values in the '*ESTIMATE*' column, making it irrelevant for the analysis. The '*AGE*' column consists of various age groups, which are only present when the '*UNIT*' column is "Crude." When the '*UNIT*' column contains age-adjusted values, the '*AGE*' column will have null values because age-adjusted statistics eliminate the effect of age distributions when comparing populations over time.

Data Preprocessing – Analyzing Trends (All Variables except *Age*) Pt. 2

```
data_new.dropna(inplace=True)
data_new.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 1050 entries, 218 to 6195
Data columns (total 16 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   INDICATOR        1050 non-null   object
 1   PANEL            1050 non-null   object
 2   PANEL_NUM        1050 non-null   int64
 3   UNIT             1050 non-null   object
 4   UNIT_NUM         1050 non-null   int64
 5   STUB_NAME        1050 non-null   object
 6   STUB_NAME_NUM    1050 non-null   int64
 7   STUB_LABEL       1050 non-null   object
 8   STUB_LABEL_NUM   1050 non-null   float64
 9   YEAR             1050 non-null   int64
 10  YEAR_NUM         1050 non-null   int64
 11  AGE_NUM          1050 non-null   float64
 12  ESTIMATE         1050 non-null   float64
 13  SEX              1050 non-null   object
 14  RACE             1050 non-null   object
 15  ETHNICITY        1050 non-null   object
dtypes: float64(3), int64(5), object(8)
memory usage: 139.5+ KB
```

The next step is to drop any rows with null values in any column, ensuring data completeness and focusing on non-null data. This resulted in a dataset of 1,050 rows, dropping 5,178 rows, since the dataset contains duplicate entries with different subcategories. This cleaned dataset can now be used for reliable and accurate analysis.

Problems Tackled

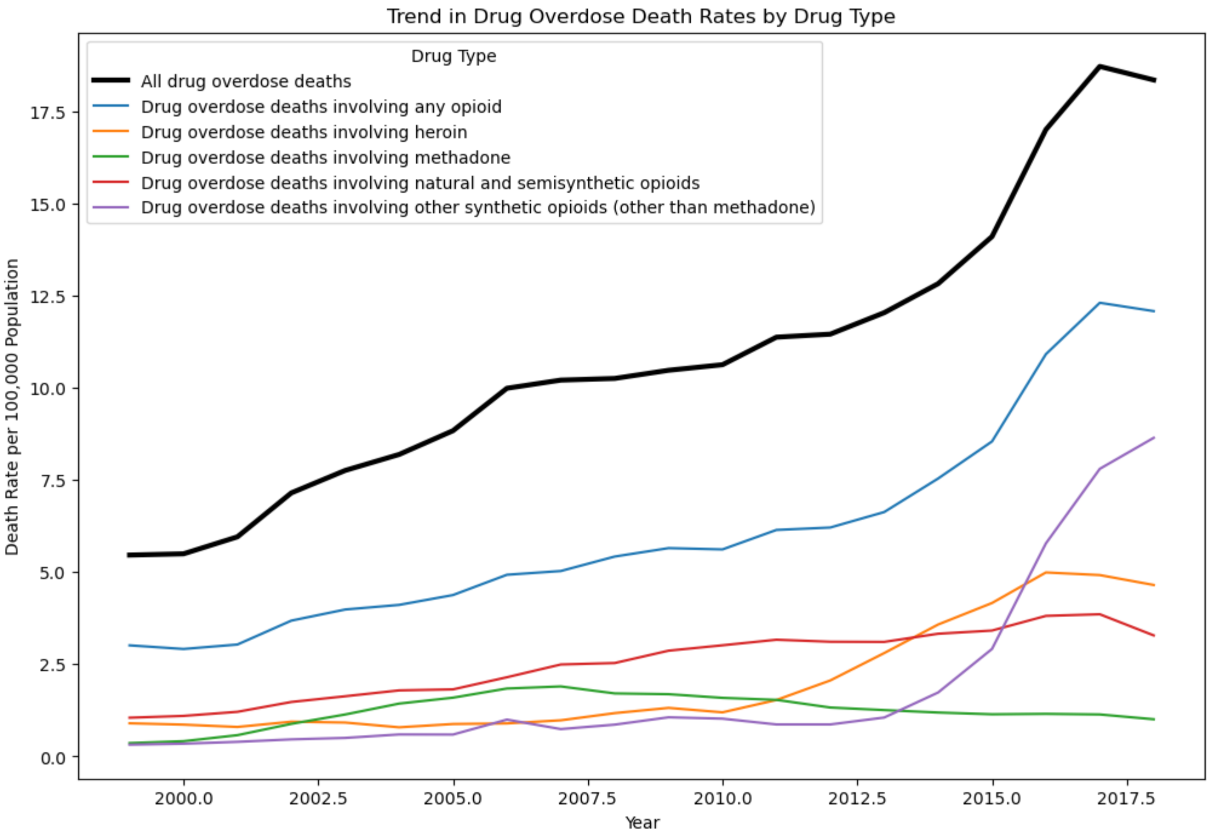
Task One	What are the trends in drug overdose death rates over time?
Task Two	Which factors are strongly associated with higher death rates?
Task Three	Major policies or interventions that have impacts on the drug overdose death rates?

Task One

What are the trends in drug overdose death rates over time?

Highlight any significant changes or patterns in the death rates over the years. Break down the trends by: Drug overdose type, Sex, Age, Race.

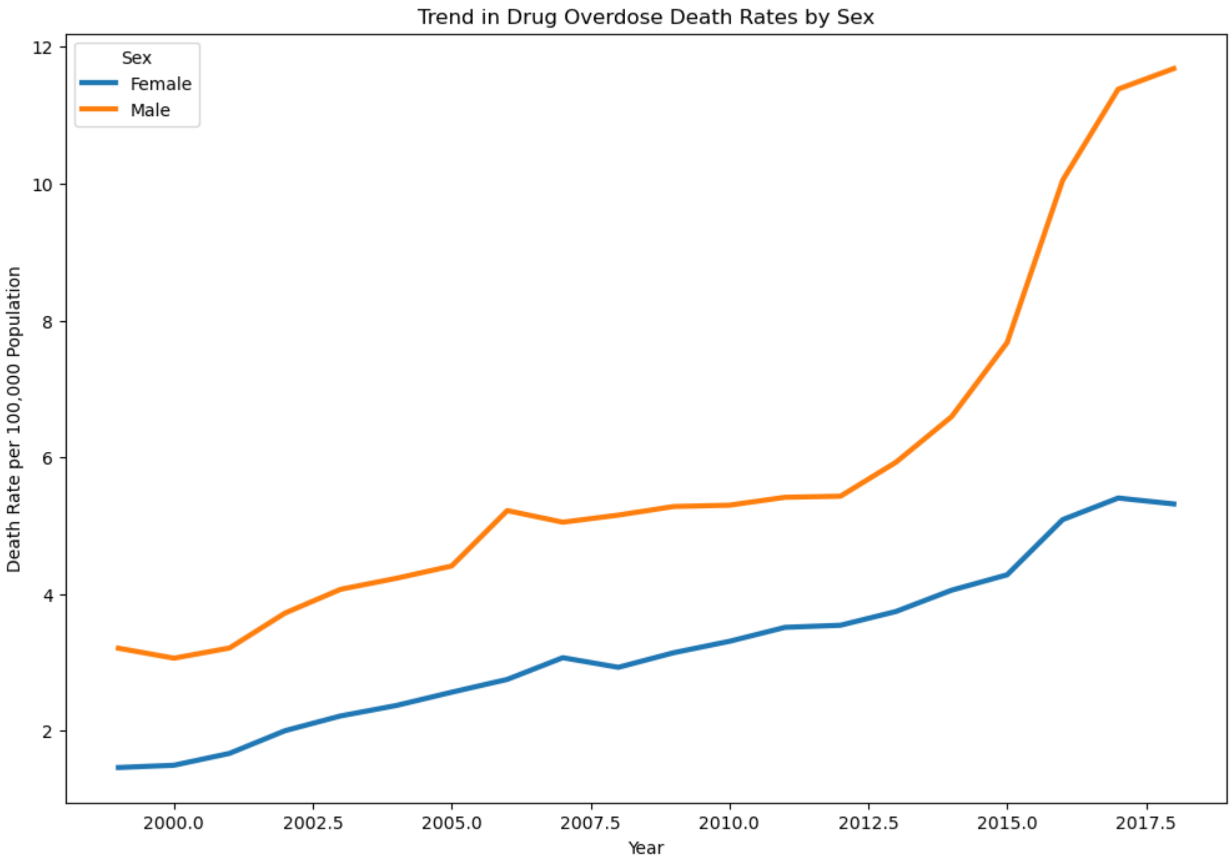
Trend by Drug Type



Information

The chart shows a sharp rise in drug overdose death rates from 1999 to 2017, with a notable spike after 2013. The overall death rate approaches 20 per 100,000 by 2017. Opioid-related deaths significantly drive this increase, with synthetic opioids other than methadone showing a dramatic surge starting in 2013. Heroin and natural and semi-synthetic opioids also rise notably. Methadone-related deaths remain stable. This data highlights the urgent need for action to address the opioid crisis.

Trend by Sex



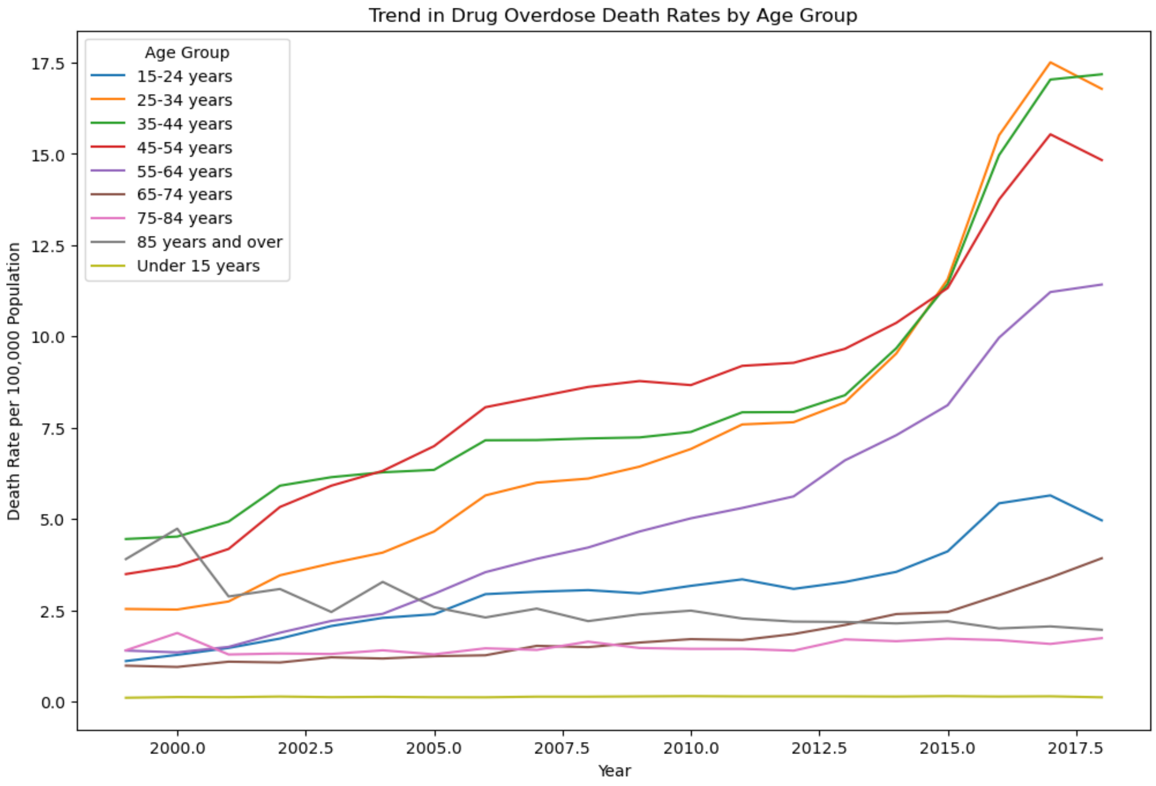
Information

The chart displays the trend in drug overdose death rates per 100,000 population from 1999 to 2017, broken down by sex. It reveals that males consistently have higher overdose death rates compared to females. While both genders show an increase over time, the rise is significantly steeper for males, especially after 2013, peaking at nearly 12 per 100,000 by 2017. In contrast, the death rate for females rises more gradually, reaching about 6 per 100,000 by the same year. This data underscores the gender disparity in drug overdose deaths, highlighting the need for targeted interventions for men who are disproportionately affected.

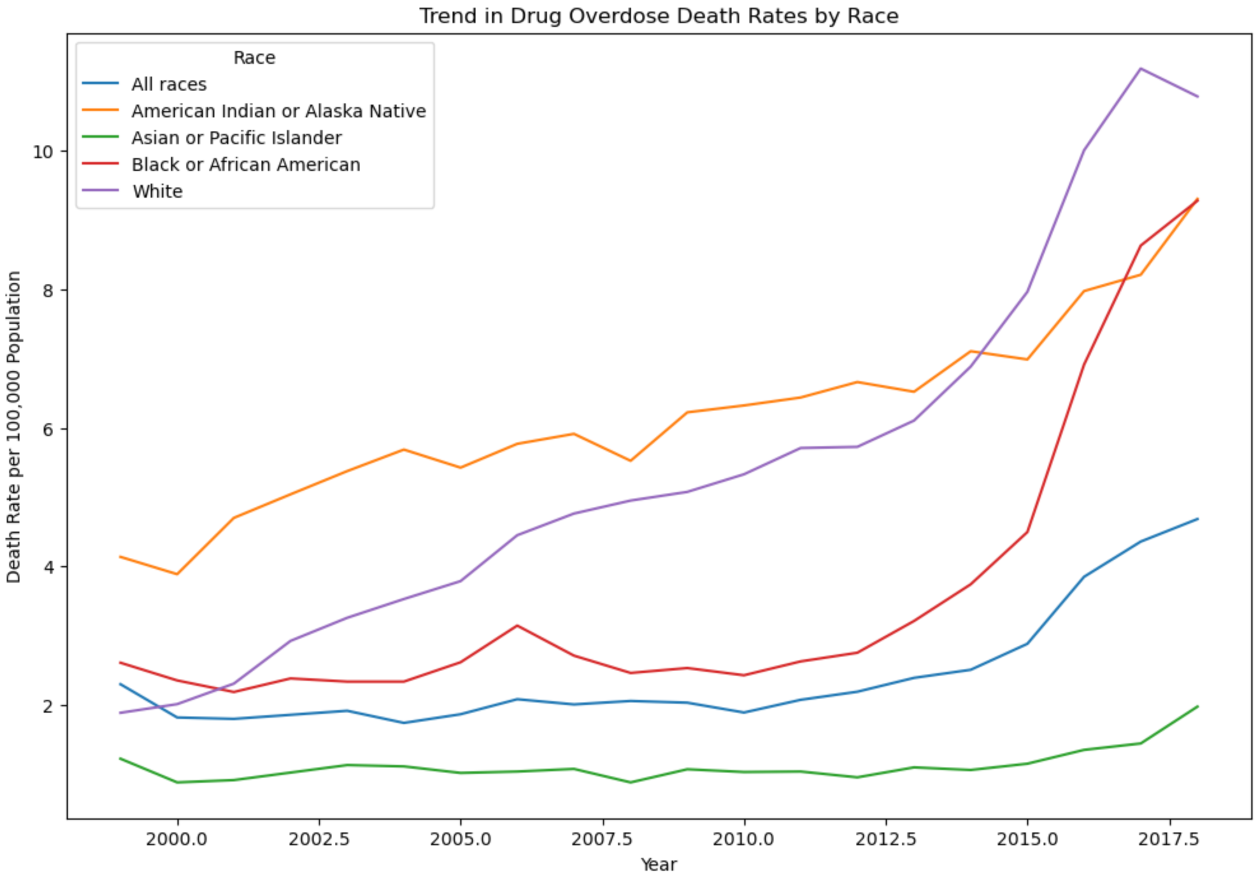
Trend by Age Group

Information

The chart illustrates the trend in drug overdose death rates per 100,000 population from 1999 to 2017, categorized by age group. It shows that the highest overdose death rates are found among those aged 25-34 years, 35-44 years, and 45-54 years, all of which have experienced a sharp increase, particularly after 2013, reaching nearly 20 per 100,000 by 2017. The 55-64 years age group also shows a significant upward trend. In contrast, the younger age groups, such as those under 15 years and 15-24 years have much lower and relatively stable death rates. Older age groups, including those over 65 years, exhibit a more gradual increase. This data highlights the age-specific nature of the drug overdose epidemic, with younger to middle-aged adults being the most affected.



Trend by Race



Information

The chart depicts the trend in drug overdose death rates per 100,000 population from 1999 to 2017, categorized by race. It reveals that white individuals experience the highest and most rapid increase in overdose death rates, especially after 2013, reaching over 15 per 100,000 by 2017. American Indian or Alaska Native individuals also show a significant rise, peaking around 10 per 100,000. Black or African American individuals and those of all races combined display notable upward trends as well. In contrast, Asian or Pacific Islander individuals maintain the lowest and relatively stable rates throughout the period. This data underscores racial disparities in the impact of the drug overdose epidemic, highlighting the disproportionate burden on white and American Indian or Alaska Native populations.

Task Two

Which factors are most strongly associated with higher death rates?

OLS Regression Model Summary

Our regression analysis examines the relationship between various factors and drug overdose death rates. The dependent variable is 'ESTIMATE' (drug overdose death rates), and the independent variables include 'PANEL', 'SEX', 'ETHNICITY', and 'RACE'.

- **R-squared: 0.474** - Approximately 47.4% of the variance in drug overdose death rates is explained by the model.
- **Adjusted R-squared: 0.465** - Adjusted for the number of predictors.
- **F-statistic: 52.56** - Indicates the model's overall significance.
- **Prob (F-statistic): 1.81e-60** - Strong evidence that the independent variables collectively influence the dependent variable.

The model is statistically significant, demonstrating that the independent variables have a significant relationship with the drug overdose death rates.

Significant Factors

Drug Type

Sex

Race

The factors we consider to calculate the significance of each variable to ***Estimate*** **(Drug Overdose Death)**

Significant Factors – Drug Type

Drug Type

Any Opioid
Coef: 4.3569, P-Value: 0.000, which refers to the significance level of 99%.
Strong positive association with death rates.

Methadone
Coef: -1.9682, P-Value: 0.000, which refers to the significance level of 99%.
Strong negative association with death rates.

The **negative coefficient for methadone-related deaths** in our regression model suggests that areas or populations with higher methadone-related deaths tend to have lower overall drug overdose deaths.

This finding could reflect the role of methadone in medication-assisted treatment programs, where it is used to manage opioid addiction under medical supervision, potentially reducing the need for and risk of more lethal opioid use. However, it is important to interpret this association with caution, as it does not imply direct causation and may be influenced by other factors in the broader context of opioid use and addiction treatment.

Significant Factors – Sex

Sex

Female
Coef: -0.3594, P-Value: 0.014, which refers to the significance level of 98.6%.
Being female is associated with lower death rates.

Male
Coef: 1.6966, P-Value: 0.000, which refers to the significance level of 99%.
Being male is associated with higher death rates.

The **negative coefficient for the variable SEX Female** in our regression model suggests that, all else being equal, being female is associated with a decrease in the estimated number of drug overdose deaths.

This finding aligns with existing statistical data showing that males generally have higher overdose death rates compared to females. Several factors may contribute to this association, including gender differences in drug use patterns, higher healthcare utilization rates among females, stronger social support networks, and biological differences in drug metabolism. It is important to note that this coefficient represents an association observed in our data and does not imply a direct causal relationship.

Significant Factors - Race

Race

American Indian or Alaska Native
Coef: 2.3205, P-Value: 0.000, which refers to the significance level of 99%.
Strong positive association with death rates.

White
Coef: 1.9881, P-Value: 0.000, which refers to the signifiance level of 99%.
Strong positive association with death rates.

Asian or Pacific Islander
Coef: -3.0819, P-Value: 0.000, which refers to the signifiance level of 99%.
Strong negative association with death rates.

The coefficients for the race variables in our regression model provide insights into the associations between race and drug overdose deaths. Specifically:

- Being **American Indian or Alaska Native** is associated with an increase in drug overdose deaths, reflecting higher rates of substance use disorders and limited access to healthcare in these communities.

- Being **Asian or Pacific Islander** is associated with a decrease in drug overdose deaths, likely due to lower rates of substance use, strong cultural norms against misuse, and robust support systems.

- Being **White** is associated with an increase in drug overdose deaths, which can be attributed to the opioid epidemic's impact, aggressive historical prescribing practices, and socioeconomic challenges.

Task Three

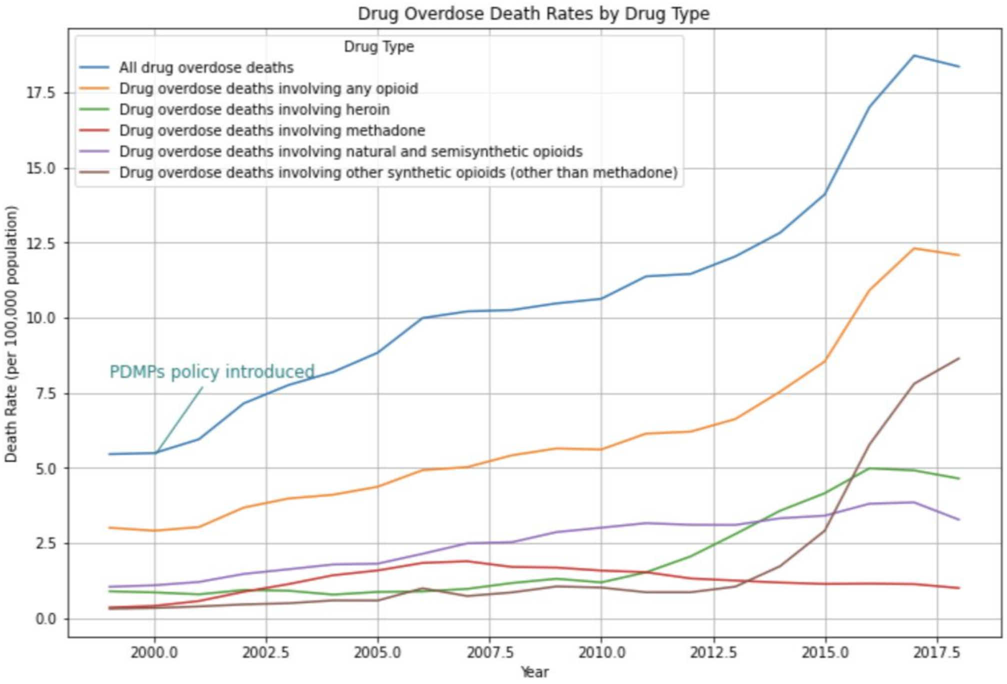
Were there any major public health policies or interventions introduced in the time period of the data?

Did they have any impact on the drug overdose death rates?

Can you identify any correlations between the introduction of specific policies and changes in death rates?

Analyze the effectiveness of these policies across different demographic groups.

Health Policy Impact - PDMP



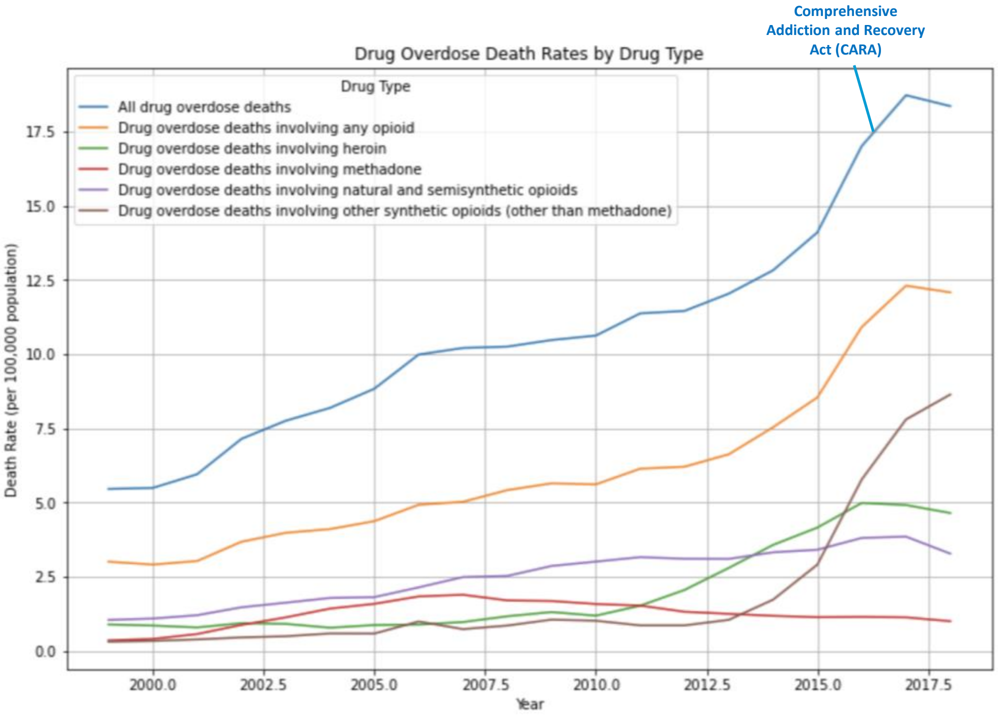
Introduction of PDMPs Policy in Early 2000s

Purpose: Prescription Drug Monitoring Programs (PDMP) are state-run programs aimed at monitoring the prescribing and dispensing of controlled prescription drugs to patients. These programs seek to identify and prevent drug abuse and diversion at the prescriber, pharmacy, and patient levels.
Initial Impact: The introduction of PDMPs in the early 2000s aimed to control prescription drug use, but the initial impact was not very pronounced.

2002-Harold Rogers Prescription Drug Monitoring Program

Purpose: In 2002, the U.S. federal government launched the Harold Rogers Prescription Drug Monitoring Program to support states in developing and implementing PDMPs.
Support: This program provided federal funding and technical assistance to states, enhancing their ability to track and control prescription drug use.
Impact: After the introduction of the Harold Rogers Program, the overall trend in drug overdose deaths (blue line) became less steep, indicating some improvement and a moderating effect of these policies.

Health Policy Impact - CARA



2013: Chinese Suppliers Turn to Fentanyl Production

- **Event:** Chinese companies began exporting large quantities of fentanyl to the U.S.
- **Impact:** Fentanyl replaced many illegal opioids due to its high potency and low cost.
- **Trend:** Rise in fentanyl-related overdose deaths (brown line).

2014-2015: Surge in Fentanyl-Related Deaths

- **Event:** Significant increase in fentanyl overdose deaths reported by states.
- **Impact:** Fentanyl's high potency and low cost rapidly replaced other illegal opioids.
- **Trend:** Steep increase in fentanyl overdose deaths.

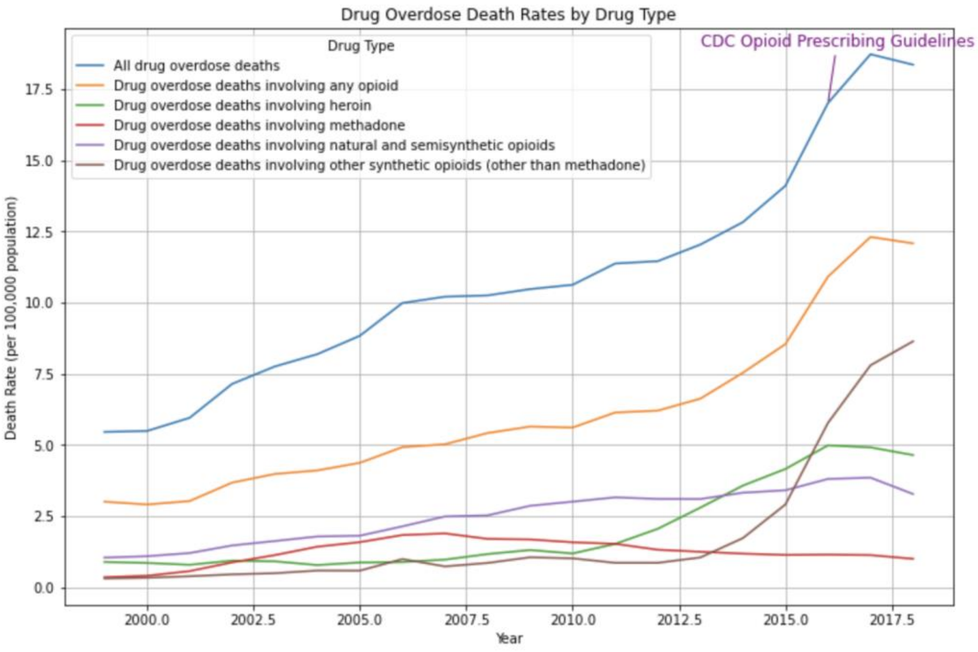
2016: Passage of the Comprehensive Addiction and Recovery Act (CARA)

- **Event:** CARA was signed into law to address the opioid crisis.
- **Impact:** Expanded naloxone access and improved treatment services.
- **Trend:** Continued rise in fentanyl-related overdose deaths.

2017: Fentanyl Scheduled as a Controlled Substance

- **Event:** Fentanyl and its analogs classified as controlled substances.
- **Impact:** Aimed to control manufacture and distribution.
- **Trend:** Slowed rise in fentanyl overdose deaths, moderating the brown line.

Health Policy Impact – CDC Opioid Prescribing Guidelines



Introduction of CDC Opioid Prescribing Guidelines (2016)

- **Event**: Introduced in 2016 to reduce inappropriate opioid prescribing.
- **Impact**: Mixed effects observed in overdose death rates.

Post-Guideline Trends

- **Overall Drug Overdose Deaths (Blue Line)**:
  Trend: Continued upward trend, but at a less steep rate.
- **Drug Overdose Deaths - Any Opioid (Orange Line)**:
  Trend: Continued rise with signs of plateauing post-2016.

Impact on Different Opioid Categories

- **Heroin (Green Line)**:
  Trend: Slight decline in overdose deaths, suggesting a reduction in heroin-related fatalities.
- **Methadone (Red Line)**:
  Trend: Stable and low overdose death rates, minimal impact from guidelines.
- **Natural and Semisynthetic Opioids (Purple Line)**:
  Trend: Slight decrease in overdose deaths, indicating a positive impact of the guidelines.
- **Other Synthetic Opioids (Brown Line)**:
  Trend: Sharp increase in overdose deaths, highlighting a shift to more dangerous synthetic alternatives like fentanyl.

Health Policy Impact –
CDC Opioid Prescribing Guidelines (Cont.)

Shift to Synthetic Opioids

- **Observation**: Significant rise in deaths involving synthetic opioids (excluding methadone) post-2016.
- **Impact**: While guidelines reduced prescription opioid misuse, they inadvertently contributed to the shift towards potent and illicit synthetic opioids.

Overall Effectiveness

- **Positive Effects**: Reduction in deaths from prescription opioids and heroin.
- **Negative Effects**: Continued rise in overall drug overdose deaths due to synthetic opioid-related fatalities.

Conclusion

- **Mixed Impact**:
  - **Positive**: Reduction in prescription opioid and heroin deaths.
  - **Negative**: Increased fatalities due to synthetic opioids.
- **Need for Comprehensive Strategies**: Regulation of prescription practices and addressing the availability/misuse of synthetic opioids are essential.

Team Data Addicts

-

[GitHub]

-

DUBSTECH

References

Anderson, I B, and T E Kearney. “Use of Methadone.” *The Western Journal of Medicine*, U.S. National Library of Medicine, Jan. 2000, www.ncbi.nlm.nih.gov/pmc/articles/PMC1070723/.

Wen, Katherine, et al. “State Policies for Prescription Drug Monitoring Programs and Adverse Opioid-Related Hospital Events.” *Medical Care*, U.S. National Library of Medicine, July 2020, www.ncbi.nlm.nih.gov/pmc/articles/PMC7985821/.