

# main

December 18, 2025

## 1 Drug Overdoes Death Rates in the United States (1999-2018)

by John Jacobsen, Marcos Negrete, Evan Vlahos, Lillie Wang

### 1.0.1 Introduction

In this project, we wanted to focus on drug overdose rates in the United States from 1999 to 2018. Drug overdose has been a major health crisis in the United States and we want to analyze how the death rates are different under different circumstances and if it is possible to predict future death rates based on previous years. To get to our results, we will be analyzing and creating different types of charts and graphs to help us better understand our focus.

### 1.0.2 Data Description

The source we will be using is from the National Center for Health Statistics {cite}cdc\_drug\_overdose\_2019 for the United States in 2019. The dataset includes death rates per 100,000 population grouped by drug types (all drugs, opioids, heroin, synthetic opioids, methadone), sex (male, female), age group (under 15, 15 - 24, 25 - 34, etc.), and lastly Race/Hispanic origin. We will be cleaning the data by filtering and removing unnecessary rows and columns.

### 1.0.3 Results

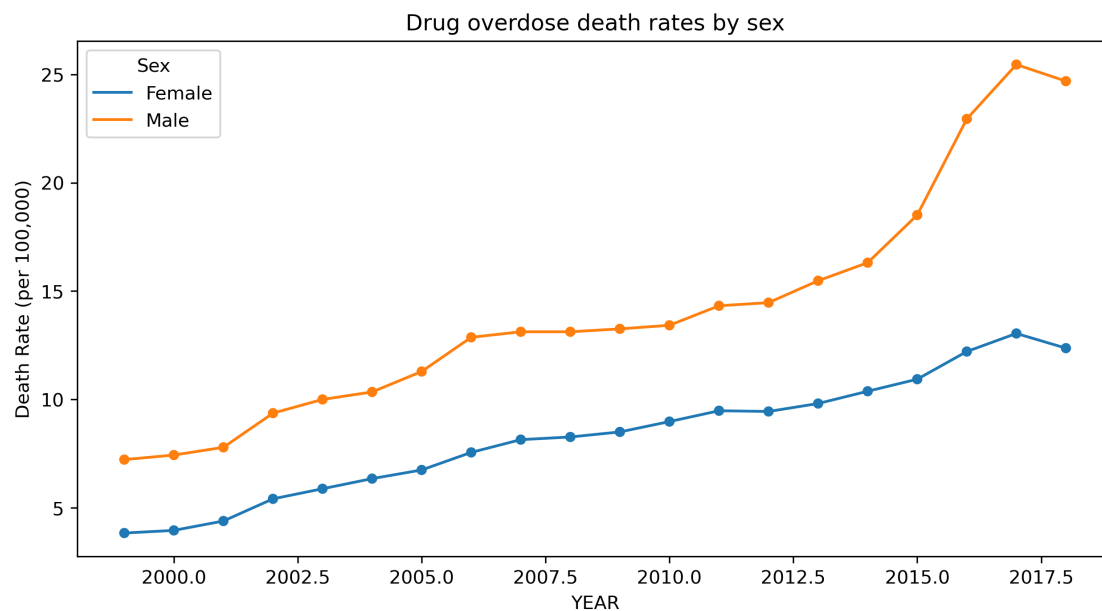
These analyses provide information about drug overdose death rates in the United States from 1999 to 2018 to see if there is any difference between groups and how it has changed over time. Using a cleaned CSV dataset and Python data science libraries, we analyzed which group had the steepest increase and which group were affected the most. The results will then help us identify trends to help us predict future death rates.

```
[3]: from IPython.display import Image
```

**Demographic Analysis** In demographic analysis, we focused on how death rates differ by sex and age groups. By sex, we see that males have a consistent higher overdose death rates than females. In 2017, the male rate was at approximately 25 per 100,000, which is double of the female rate at approximately 13. By age group, the highest death rates are concentrated in the middle-aged adults, around 25-55 years old. In 2017, it showed the highest spike while other groups like 25 and under, 65 and over stayed under 5 death rates per 100,00. The groups that showed the highest absolute increase are males aged 25-34, with 38.1 deaths per 100,000. The highest percent increase group are the 55-64 year olds with a massive spike of 659.2% for male and 468.6% for females.

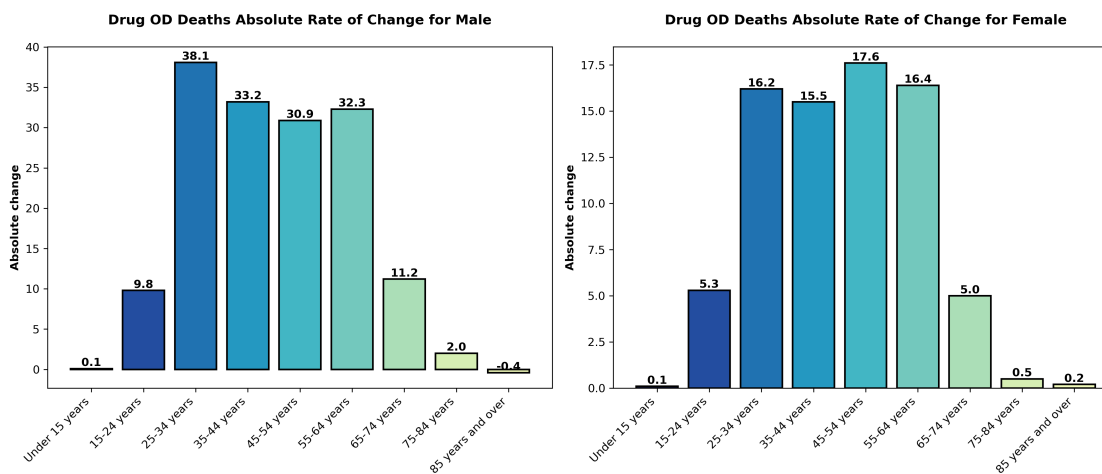
```
[4]: Image("figures/drug_overdose_death_rates_by_sex.png")
```

[4]:



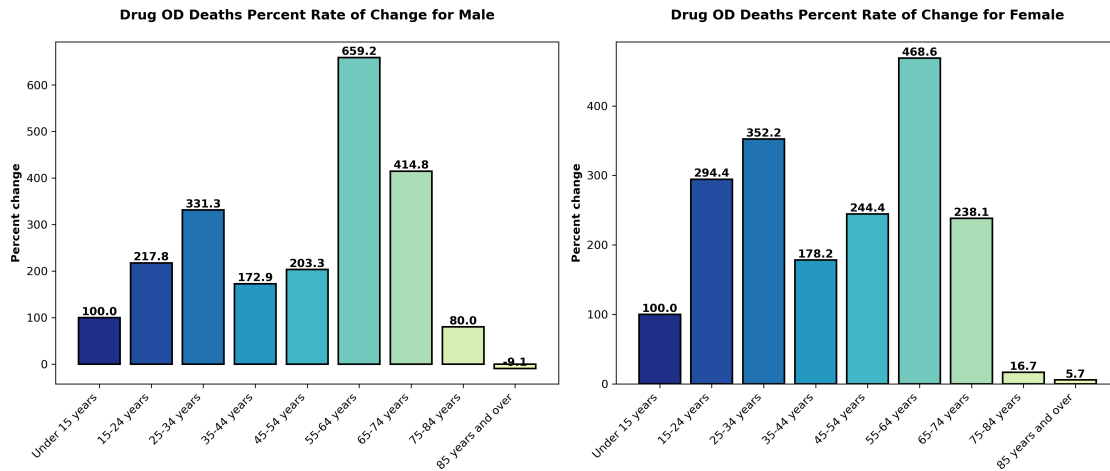
```
[5]: Image("figures/drug_overdose_deaths_absolute_change_by_sex.png")
```

[5]:

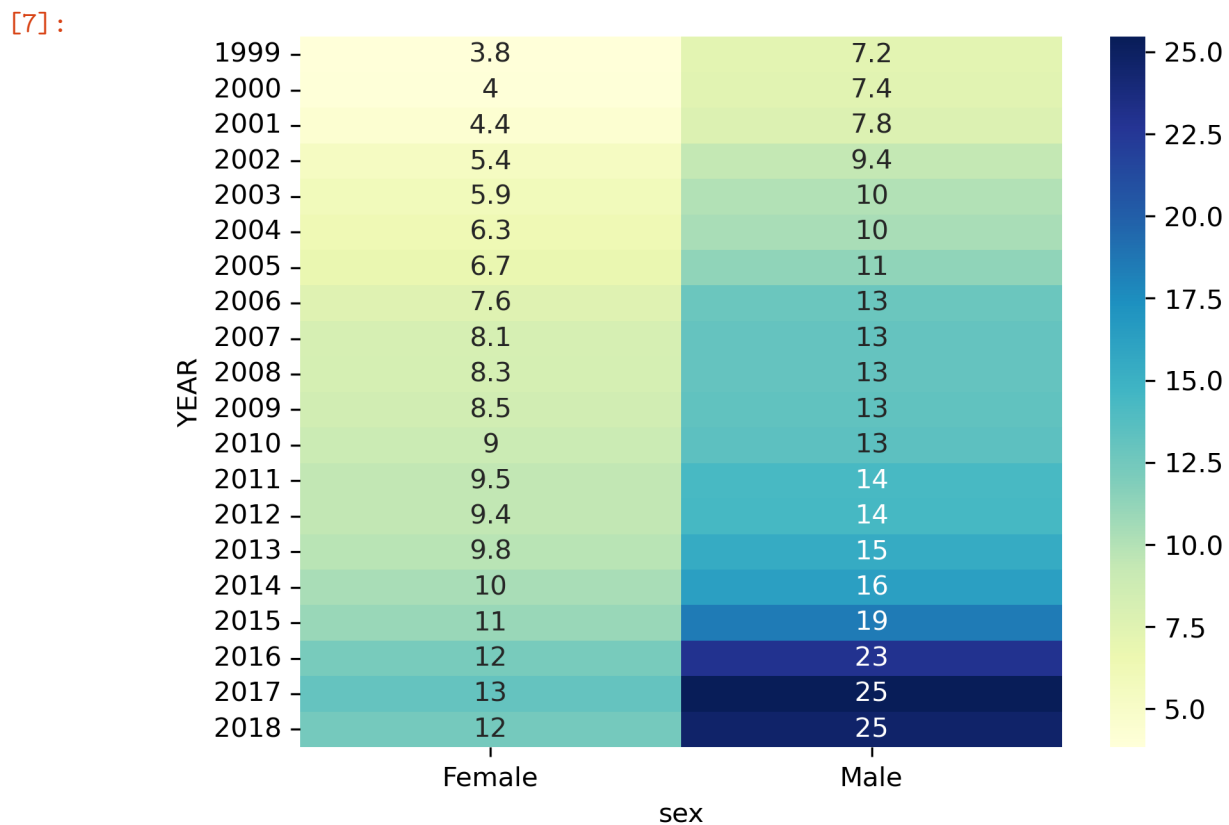


```
[6]: Image("figures/drug_overdose_deaths_percent_change_by_sex.png")
```

[6]:

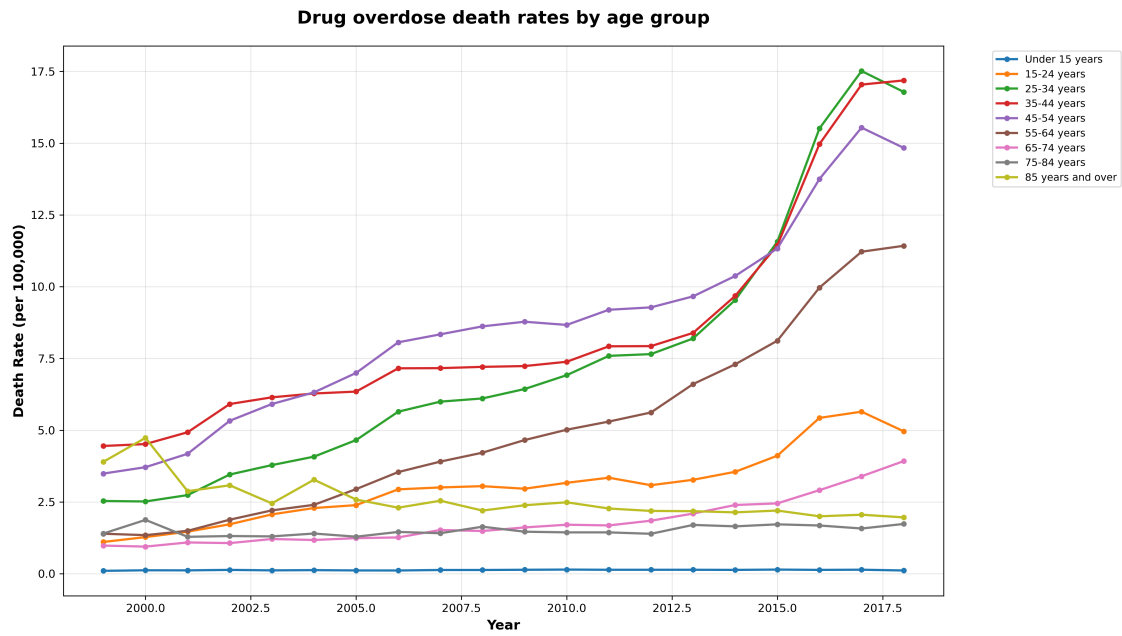


```
[7]: Image("figures/heatmap_by_sex.png")
```



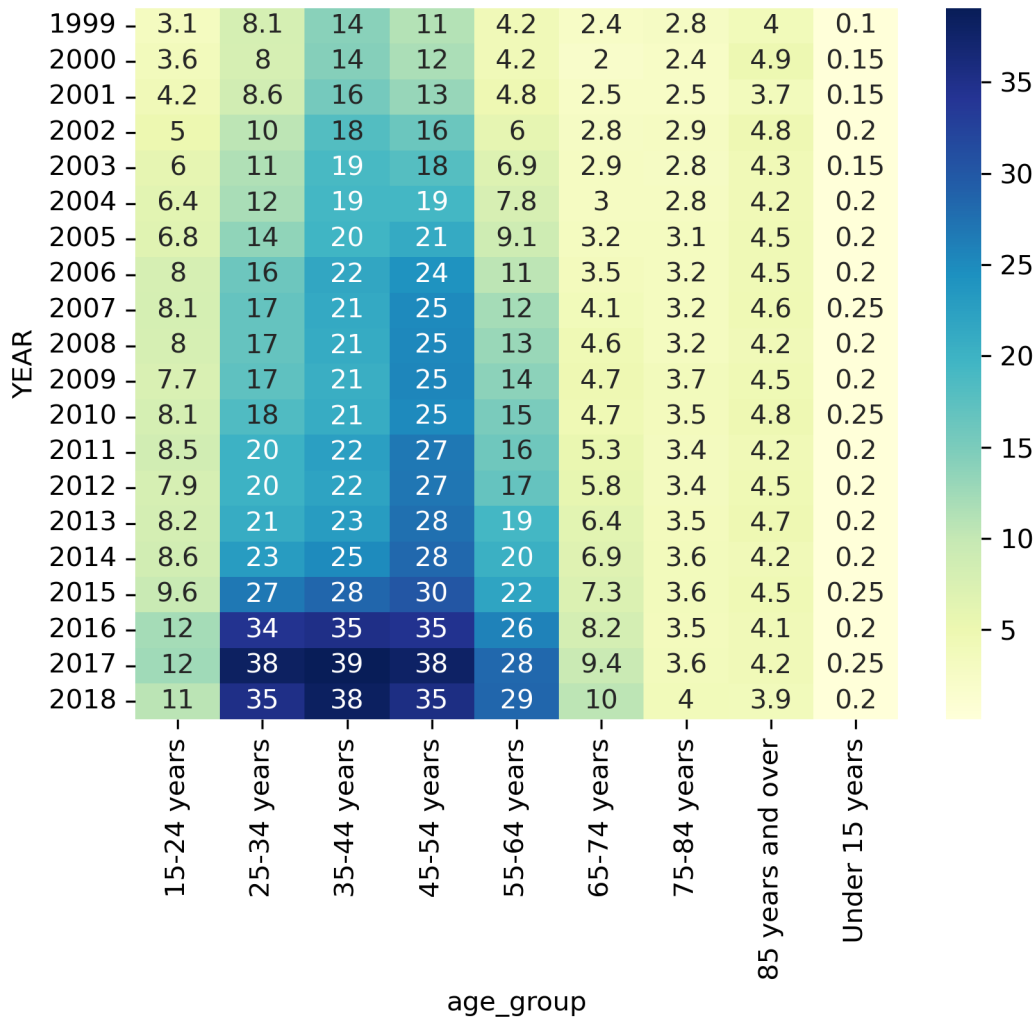
```
[8]: Image("figures/drug_overdose_death_rates_by_age_group.png")
```

[8]:



```
[9]: Image("figures/heatmap_by_age_group.png")
```

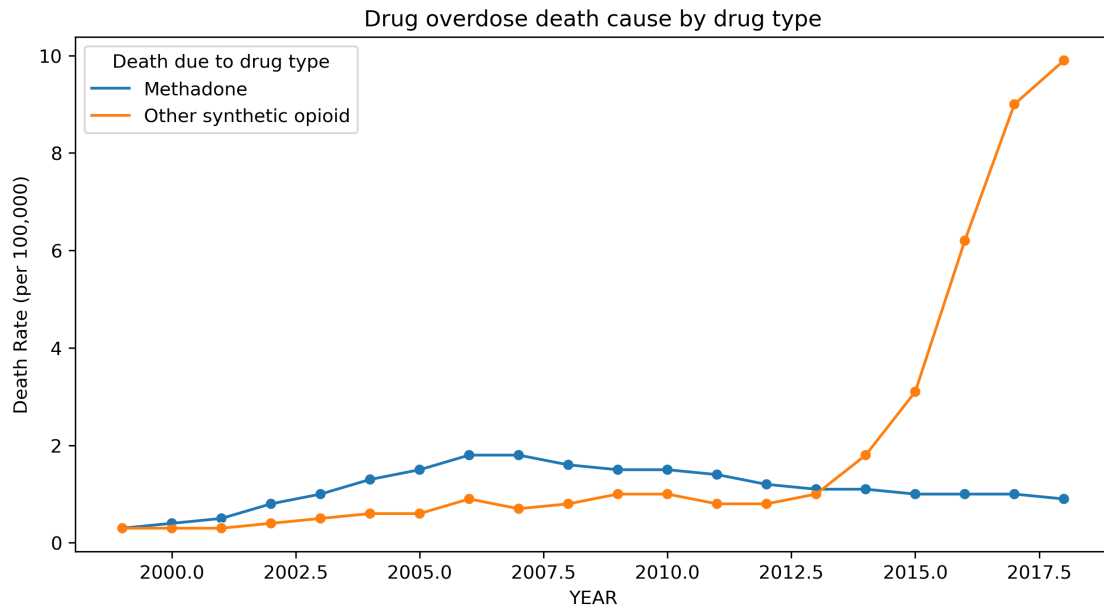
```
[9]:
```



**Opioid Analysis** In opioid analysis, we focused on the proportion of synthetic opioid deaths over time and how racial groups are affected. From the graph, we can see that since 2013, synthetic opioids have become the primary driver of overdose deaths. In 2018, the death rates involving synthetic opioids was nearly 70% compared to the 10% in the early 2000s. While methadone had rises from 2002 to 2010 but went back down over the years. The racial group that showed the highest ratio, and affected the most during 2010 and 2018 was Asian or Pacific Islander at roughly 60%. While the other groups stayed around and under 50%.

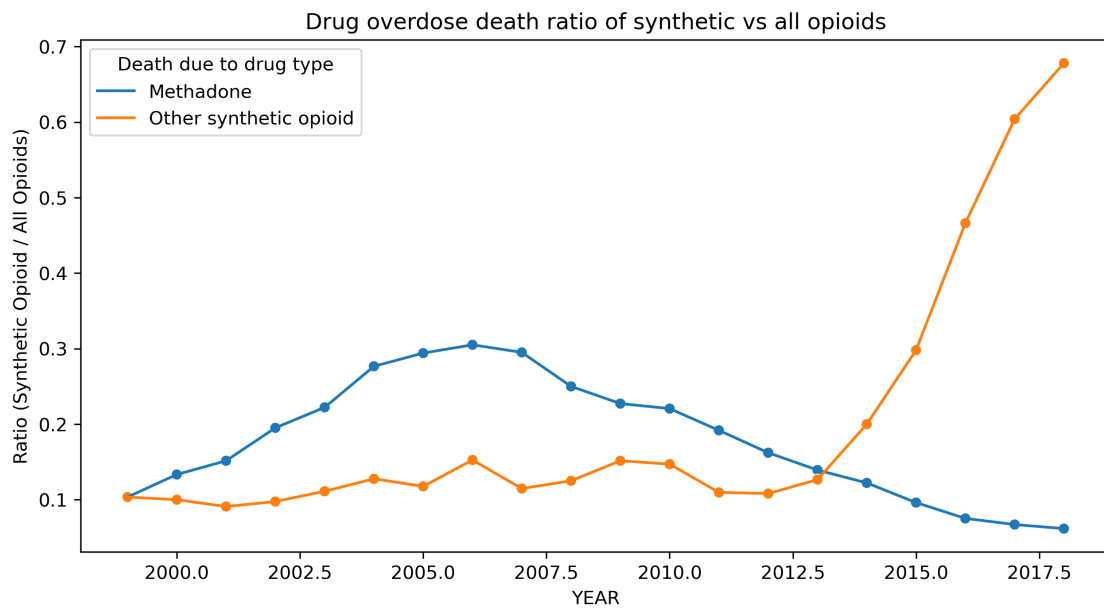
```
[10]: Image("figures/drug_overdose_death_cause_by_drug_type.png")
```

```
[10]:
```



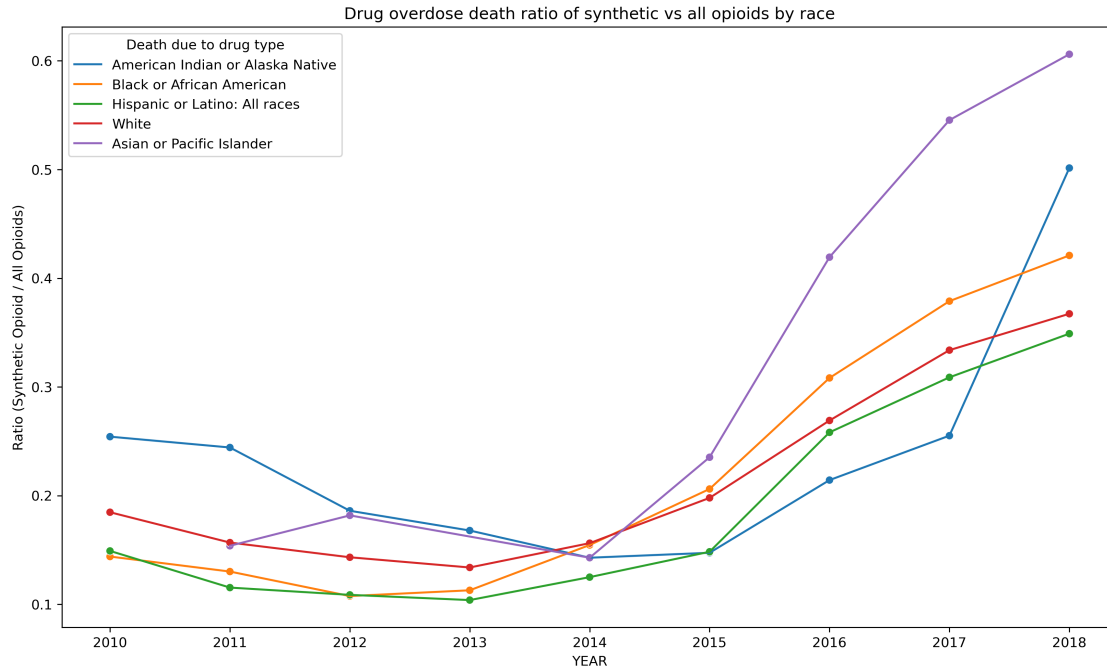
```
[11]: Image("figures/drug_overdose_death_ratio_opioids.png")
```

[11]:



```
[12]: Image("figures/drug_overdose_death_ratio_opioids_by_race.png")
```

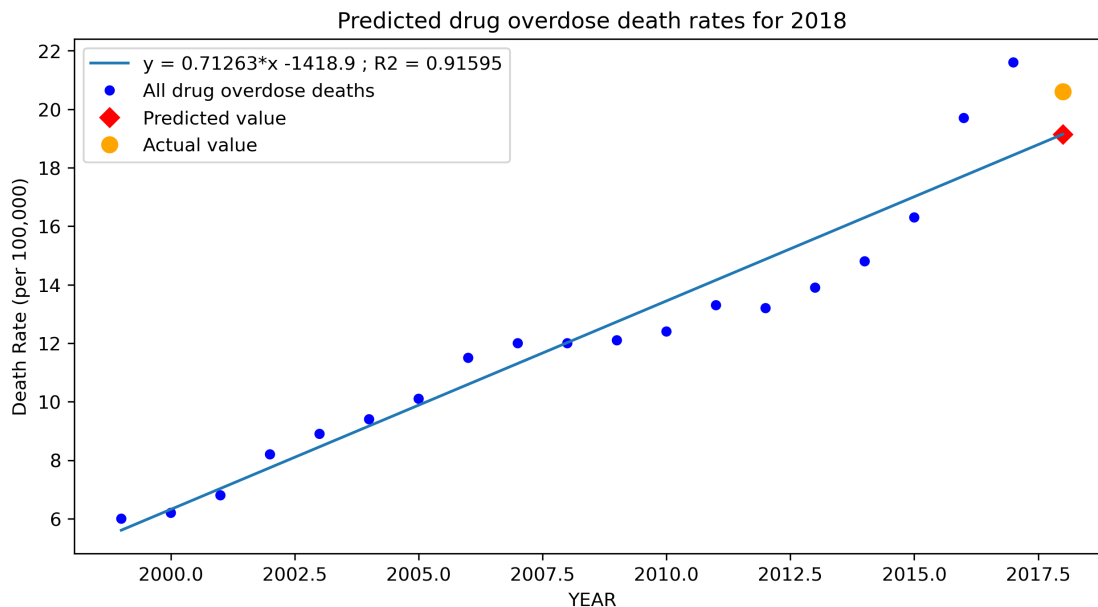
[12]:



**Trend Prediction** In trend prediction, we mainly wanted to figure out if we could use previous data from 1999 to 2017 to predict 2018 death rates. Based on the graph, it is possible to predict 2018 death rates using linear regression, however there is a notable gap between the predicted and actual values.

[13]: `Image("figures/drug_overdose_prediction.png")`

[13]:



Even though we were able to get many clear graphs and results from the provided data, there are still limitations to our project.

Overlapping drug categories: Many overdose deaths involve mixing drugs, and because the data counts these drugs separately, it is hard to know which drug was the primary cause of death.

Missing standard errors: The graphs show the average death rates and not the standard errors, and without them, it is harder to prove every small change is a real trend.

Ecological data: This project studies large groups of people like race and age groups, meaning that we see trends for the group as a whole, so we can not make assumptions about any specific individual.

#### 1.0.4 Conclusion

This project shows that drug overdose crisis has changed significantly over the past decades, becoming more dangerous due to the rise of synthetic opioids. While men consistently have higher death rates than women, the crisis has expanded to middle-aged adults. In recent years, synthetic opioids were involved in the majority of deaths across all groups, particularly affecting the Asian or Pacific Islanders. Lastly, our predictions showed that the crisis is accelerating so quick that linear regression is underestimating the overdose death rate.

#### 1.0.5 Author Contributions

**John Jacobsen:** Code infrastructure and testing frameworks including creation of Python module (`src/analysis_utils.py`) with utility functions for demographic label parsing, data filtering, rate change calculations, and trend modeling. Developed comprehensive testing suite (`tests/test_analysis_utils.py`) with unit tests for all functions. Set up repository infrastructure including `environment.yml`, `Makefile`, `.gitignore`, and `LICENSE` file. Managed build automation and code quality standards.

**Marcos Negrete:** Data acquisition, cleaning, and exploration. Downloaded and documented CDC drug overdose death data from Data.gov {cite}`data_gov_overdose`. Created data cleaning pipeline in `LoadOverdoseData.ipynb` to fix data inconsistencies, parse demographic labels, filter unreliable estimates, and export processed datasets. Developed exploratory data analysis in `EDA.ipynb` with summary statistics, trend visualizations, and demographic comparisons. Authored data documentation and README files.

**Evan Vlahos:** Analysis notebooks `DemographicAnalysis.ipynb`, `OpioidAnalysis.ipynb`, and `TrendPrediction.ipynb`. Minor edits to `LoadOverdoseData.ipynb` to save data to CSV files in the `data/` folder. Generating PNG images of the graphs in the `figures/` folder.

#### **Lillie Wang:**

Technical writing: Contributed to main narrative notebook, wrote clear introductions/conclusions, and synthesizing findings from multiple analysis notebooks into coherent storylines

MyST configuration: Initialize `myst.yml`, configuring GitHub Pages deployment, and troubleshooting website build issues



Documentation creation: Setup README files, structuring author contributions, and creating clear installation/usage instructions

Integration tasks: Initialize Binder setup and ensuring all project components work together seamlessly for reproducibility

### 1.0.6 References

```
@misc{CDC_DrugOverdoseDeathRates_95ax-ymtc,  
  author      = {{Centers for Disease Control and Prevention (CDC)}},  
  title       = {Drug overdose death rates, by drug type, sex, age, race, and Hispanic origin},  
  year        = {2025},  
  howpublished = {\url{https://data.cdc.gov/api/views/95ax-ymtc/rows.csv?accessType=DOWNLOAD}},  
  note        = {Public dataset. Accessed: YYYY-MM-DD},  
  institution = {U.S. Centers for Disease Control and Prevention},  
  type        = {Data set}  
}
```