

# **TEAM HIRE ME NOW**

---

**Health: Drug Overdose in  
USA**



# MEET THE TEAM



DIANE HOANG

**Project & Data Cleaning  
Team Lead**

<https://www.linkedin.com/in/dianehoang/>



AARON THAI

**Data Visualization  
Team Lead**

<https://www.linkedin.com/in/aaron-thai/>



DANA ABDIRAKHYM

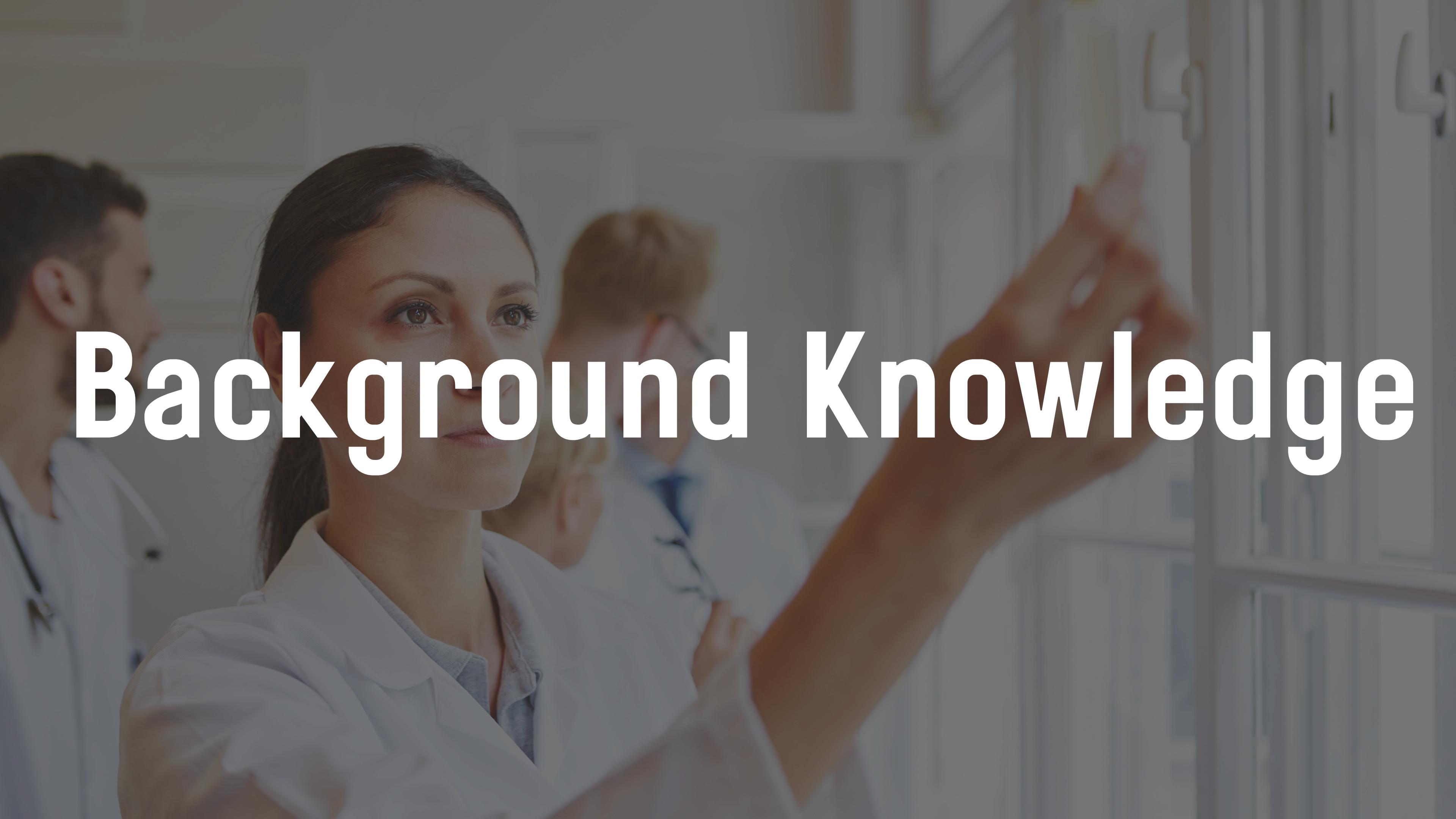
**Machine Learning  
Team Lead**

<https://www.linkedin.com/in/danabdirakhym/>

# Agenda

1. Background Knowledge
2. Data Cleaning
3. Data Visualization
4. Machine Learning

# Background Knowledge



# DRUG OVERDOSES

Drug overdoses are **extremely dangerous.**

Different drug types have different lethaliites.

Among the most dangerous types of drugs are **Synthetic Opioids**, some of the most widespread being Fentanyl and Carfentanil, which is **100x more lethal** than Fentanyl,

and **5000x** more lethal than **Heroin**.

**Source:** <https://www.dea.gov/press-releases/2016/09/22/dea-issues-carfentanil-warning-police-and-public>

# DATA CLEANING



## UNIT

Deaths per 100,000 resident population, age-adjusted, Deaths per 100,000 resident population, crude



Renamed Unique Values

Age-adjusted Crude

## PANEL

All drug overdose deaths, Drug overdose deaths involving any opioid, Drug overdose deaths involving natural and semisynthetic opioids, Drug overdose deaths involving methadone, Drug overdose deaths involving other synthetic opioids (other than methadone), Drug overdose deaths involving heroin

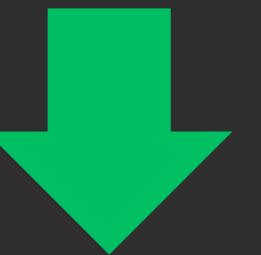


Renamed Unique Values

All ODS  
Any Opioids  
Nat. & Semi-opioids  
Methadone  
Other  
Synthetics(excl.Met hadone)  
Heroin

## STUB\_NAME

Total, Sex, Sex and race, Sex and race and Hispanic origin, Age, Sex and age, Sex and race (single race), Sex and race and Hispanic origin (single race)



Made new columns: Sex, Age, Race, or Hispanic Origin column

Sex  
Age  
Race  
Hispanic Origin

## STUB\_LABEL

All persons, Male, Female, Male: White, Male: Black or African American, Male: American Indian or Alaska Native, Male: Asian or Pacific Islander, Female: White, Female: Black or African American, Female: American Indian or Alaska Native, Female: Asian or Pacific Islander, Male: Hispanic or Latino: All races, Male: Not Hispanic or Latino: White, Male: Not Hispanic or Latino: Black, Male: Not Hispanic or Latino: American Indian or Alaska Native, Male: Not Hispanic or Latino: Asian or Pacific Islander, Female: Hispanic or Latino: All races, Female: Not Hispanic or Latino: White, Female: Not Hispanic or Latino: Black, Female: Not Hispanic or Latino: American Indian or Alaska Native, Female: Not Hispanic or Latino: Asian or Pacific Islander, Under 15 years, 15-24 years, 25-34 years, 35-44 years, 45-54 years, 55-64 years, 65-74 years, 75-84 years, 85 years and over, Male: Under 15 years, Male: 15-24 years, Male: 25-34 years, Male: 35-44 years, Male: 45-54 years, Male: 55-64 years, Male: 65-74 years, Male: 75-84 years, Male: 85 years and over, Female: Under 15 years, Female: 15-24 years, Female: 25-34 years, Female: 35-44 years, Female: 45-54 years, Female: 55-64 years, Female: 65-74 years, Female: 75-84 years, Female: 85 years and over, Male: Not Hispanic or Latino: Asian, Male: Not Hispanic or Latino: Native Hawaiian or Other Pacific Islander, Female: Not Hispanic or Latino: Asian, Female: Not Hispanic or Latino: Native Hawaiian or Other Pacific Islander



Sorted into either Sex, Age, Race, or Hispanic Origin column

If the data didn't fit into one of the columns it would be left as non applicable, Not Hispanic.

A close-up photograph of a person's fingers holding a small, rectangular wooden block. The block has a light-colored wood grain texture. In large, bold, black capital letters, the word "QUALITY" is printed across the top half. Below it, in smaller black capital letters, is the phrase "IS THE BEST BUSINESS PLAN". In the bottom half of the block, the words "DATA VISUALIZATION" are printed in large, white, sans-serif capital letters.

**QUALITY**  
**IS THE BEST**  
**BUSINESS**  
**PLAN**

**DATA VISUALIZATION**

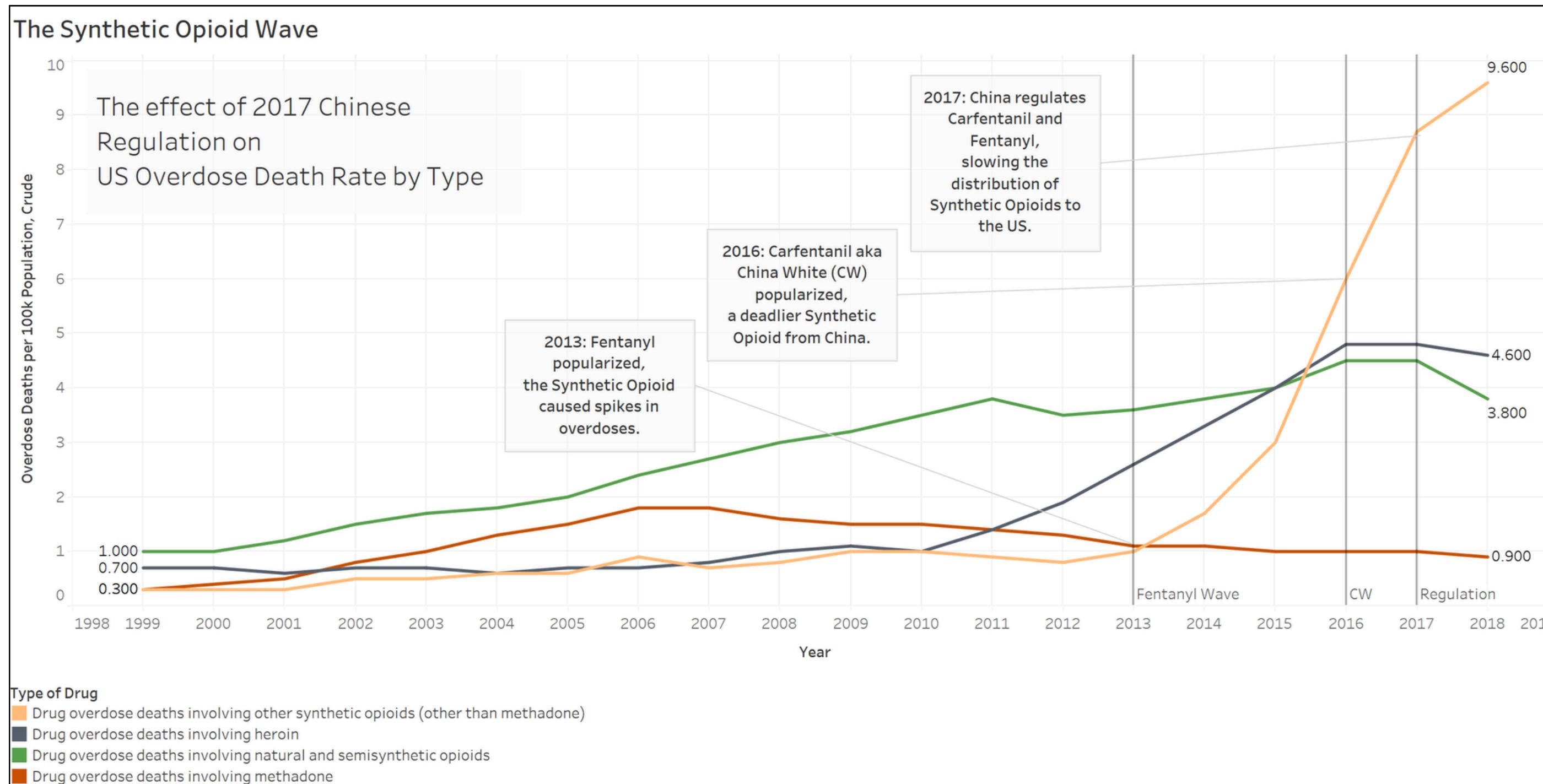
# PROMPT

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

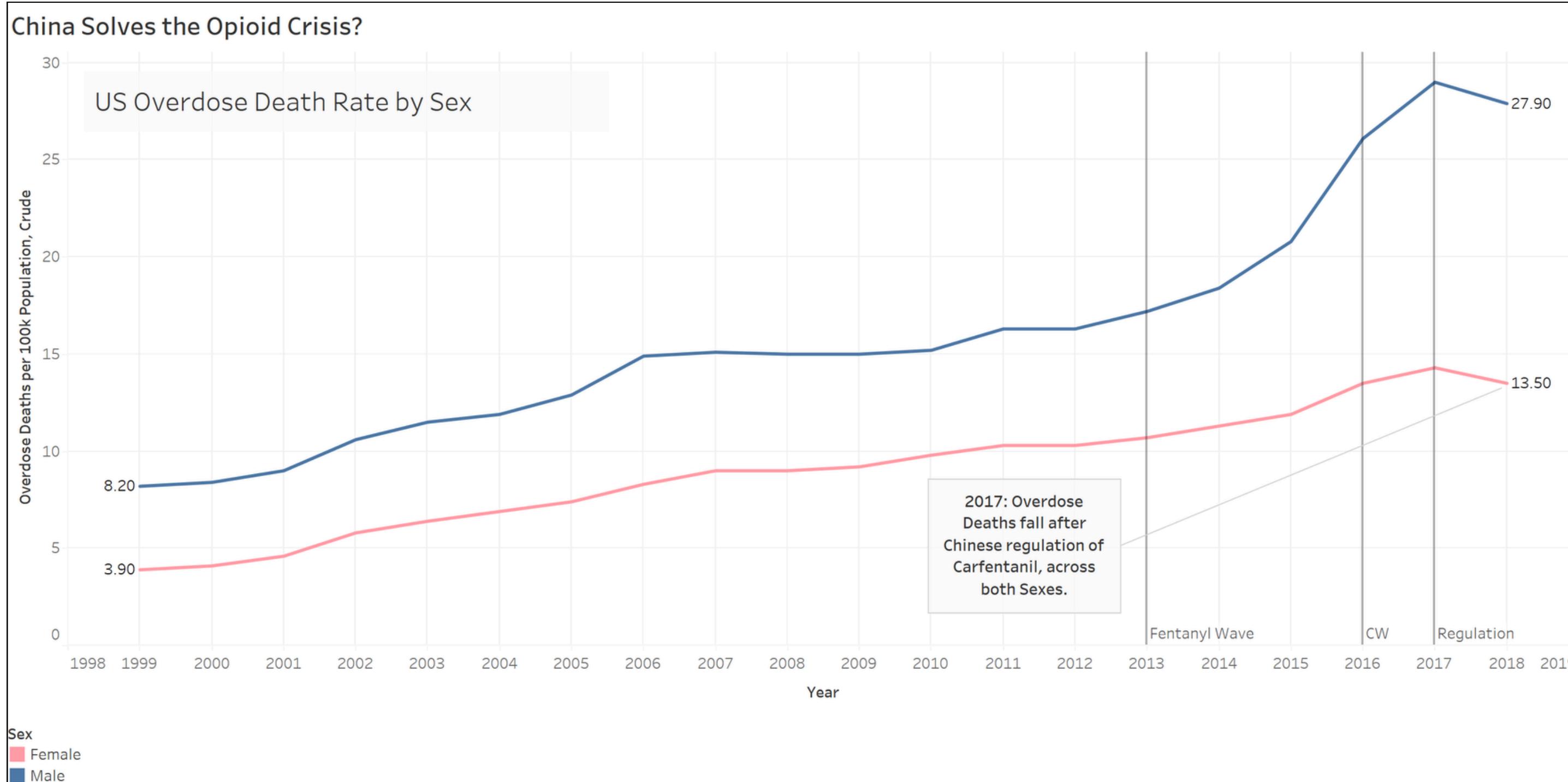
# Were there any major public health policies or interventions introduced in the time period of the data?

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



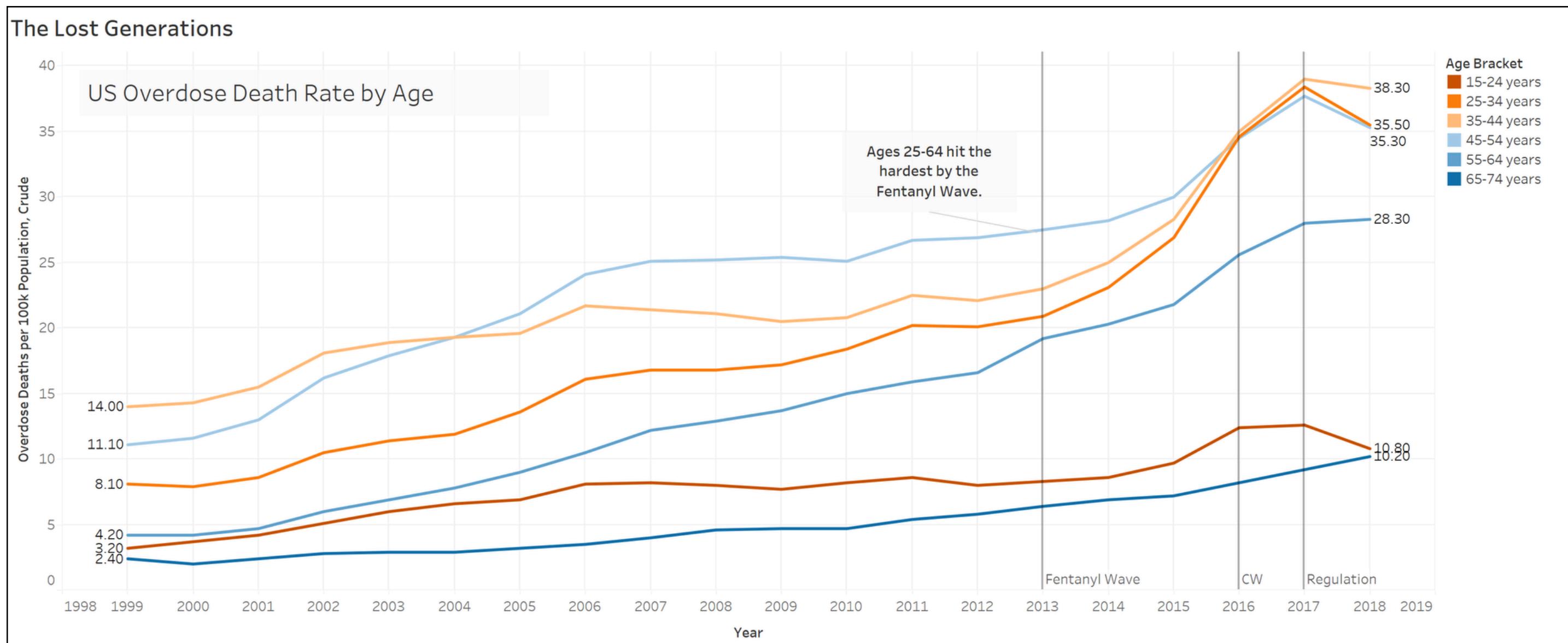
In 2017, China regulated Carfentanil, slowing US drug overdose death rates.

# Analyze the effectiveness of these policies across different demographic groups.



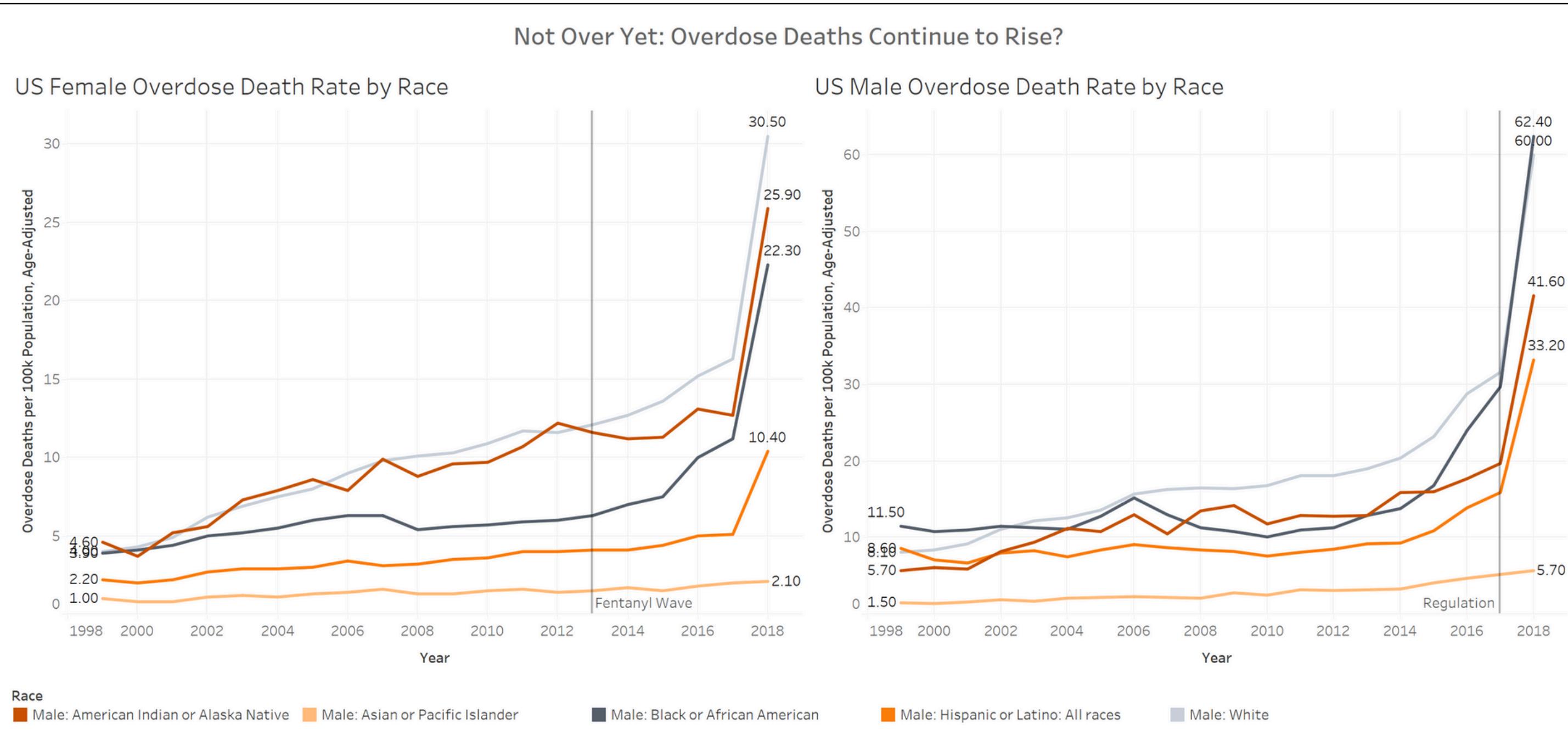
The Chinese regulation slowed US drug overdose death rates across both sexes.

# 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... Age



Drug overdose death rates fell for age ranges 15-54 years in 2017, but continued rising for age ranges 55-74 years.

# Break down the trends by... Race



Drug overdose death rates seemed to continue **rising** across all races for both sexes.

# Machine Learning



# PROMPT

Build a machine learning model to identify individuals or demographics at the highest risk of drug overdose deaths. Use features like drug type, age, sex, race, and Hispanic origin.



# Estimation Task

Estimation involves predicting the value of a continuous target variable based on input features. It aims to estimate the relationship between the input variables and the target variable.

**Example: Linear regression, polynomial regression, support vector machines.**



**Build a machine learning model to identify individuals or demographics at the highest risk of drug overdose deaths.**

7 Models used to predict “Estimate”

## Random Forest Regressor

high prediction accuracy:

The RMSE of 0.3903 suggests predictions are very close to the actual values of ESTIMATE, with an average error of less than half a unit.

high explanatory power:

The R<sup>2</sup> value of 0.9963 signifies that the model has a very strong explanatory power, as it can account for 99.6% of the variability in the ESTIMATE values based on the input features

Metric	Training Set	Validation Set
RMSE	0.3903	0.3903
R <sup>2</sup>	0.9963	0.9963

# Top 8 features predicting the drug overdose death

Rank	Feature	Importance
1	PANEL_NUM	0.330176
2	AGE_NUM	0.269693
3	YEAR_NUM	0.176921
4	STUB_LABEL_NUM	0.155854
5	Sex_Male	0.053692
6	STUB_NAME_NUM	0.009756
7	Race_Hispanic_or_Latino	0.003539
8	UNIT_NUM	0.000369

- **PANEL\_NUM, AGE\_NUM, YEAR\_NUM, STUB\_LABEL\_NUM:** These numeric features are the most influential in predicting the ESTIMATE.
- **Sex\_Male and Race\_Hispanic or Latino** These categorical features also play significant roles in the predictions.

Link to the ML Model:  
<https://colab.research.google.com/drive/1veNK0xzmQb6tf645iwxYOo4OzSoG0uBC?usp=sharing>



**Are there any  
questions?**

---

**THANK YOU!**