

Drugs and Jobs: The effect of unemployment on drug overdose deaths in America

Evan Arnold and Caleb Ren

12/06/2019

Introduction and Motivation

Overdose deaths in the US have increased dramatically since 1999

Exploratory Data Analysis Methods

Data Summary

As a next step in our project, we collected data from the CDC in the form of the Vital Statistics Rapid Release dataset (VSRR). The VSRR data contains provisional counts of drug overdose deaths in the US as reported by agencies from all 50 states and the District of Columbia. The data is collected on a monthly basis.

The data of interest to this project is the number of deaths in each state as a result of drug overdose. Drug overdoses are counted by state agencies in accordance to World Health Organization standards, which lay out the basic guides for reporting agencies to code and classify causes of death. Drug categories that are represented in this dataset include the major drivers of the opioid epidemic like heroin (coded by T40.1), natural opioid analgesics (morphine and codeine), synthetic opioids (oxycodone, hydrocodone, oxycodone; T40.2), methadone (T40.3), other synthetics (fentanyl, tramadol; T40.4) and other drugs like cocaine, methamphetamine, etc.

There were over 26052 data points from the VSRR dataset. Of those data points, many are individual observations of different coded deaths from different drugs; after reshaping and data cleaning, there are now 2652 individual observations. The data ranges from 2015 to 2019, with each state reporting 52 observations (once per month). Overdose deaths range from 55 deaths in the month of May 2015 in South Dakota to a high of 5697 in Pennsylvania in September of 2017.

Unemployment data was sourced from the Bureau of Labor Statistics. Unemployment data is published in monthly increments from the Bureau of Labor Statistics by state. Data is published beginning in 1976 and is published on the first of each month describing the previous month's unemployment rate.

There is a very specific definition of who in the labor force is considered *unemployed*. According to the BLS, those who are currently unemployed are those who are "jobless, looking for a job, and available for work." People who are incarcerated, in a nursing home, or in a mental health care facility are not considered unemployed as they are not fit for work.

Using this definition, data was scraped from the BLS website and aggregated by each state and the District of Columbia. The unemployment rate in percent is given by the `unemployment` column. The lowest unemployment rate in a given state and month is Vermont in 2019 with a 2.1% unemployment rate. The highest rate is DC in 2015 with a 7.4% unemployment rate. The data itself is roughly Normally distributed with a mean of 4.2% and a median of 4.31%.

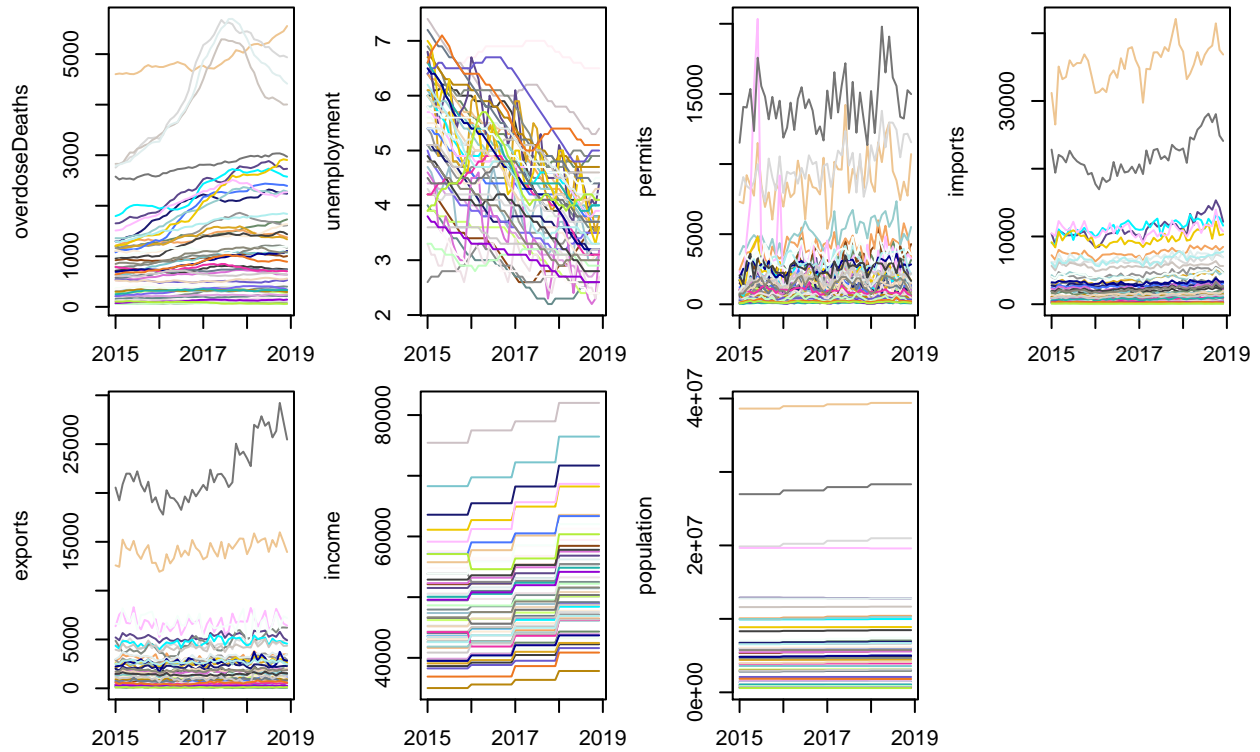
St. Louis Datasets permits: housing units authorized by building permits (raw count). This is a proxy for housing development. imports: imports in millions of dollars. This is a proxy for in-state manufacturing. income: annual income per capita

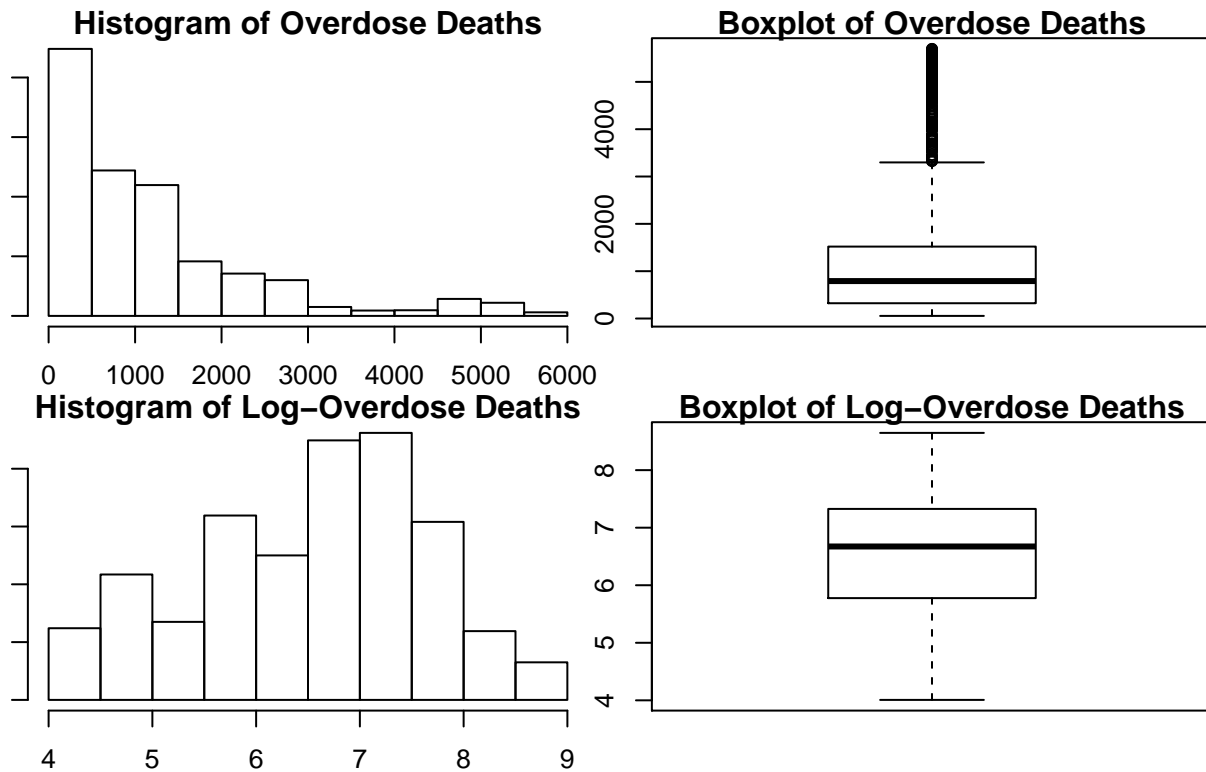
Census Bureau Dataset population: raw population

NOTE: For each state, we use the population estimate for the previous year for the entire year. This is due to a lack of available data as well as slow population growth across states.

R Data region: data from R. We have to include the region for DC specifically

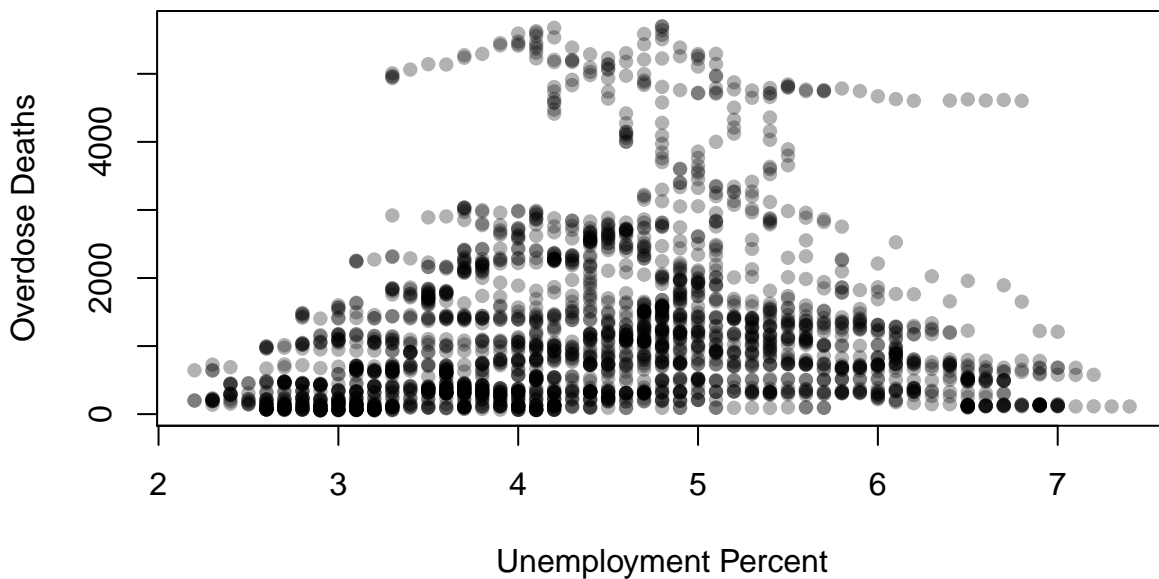
EDA





We see that the data is much closer to a Normal distribution if we apply a log transformation.

Overdose Deaths vs. Unemployment



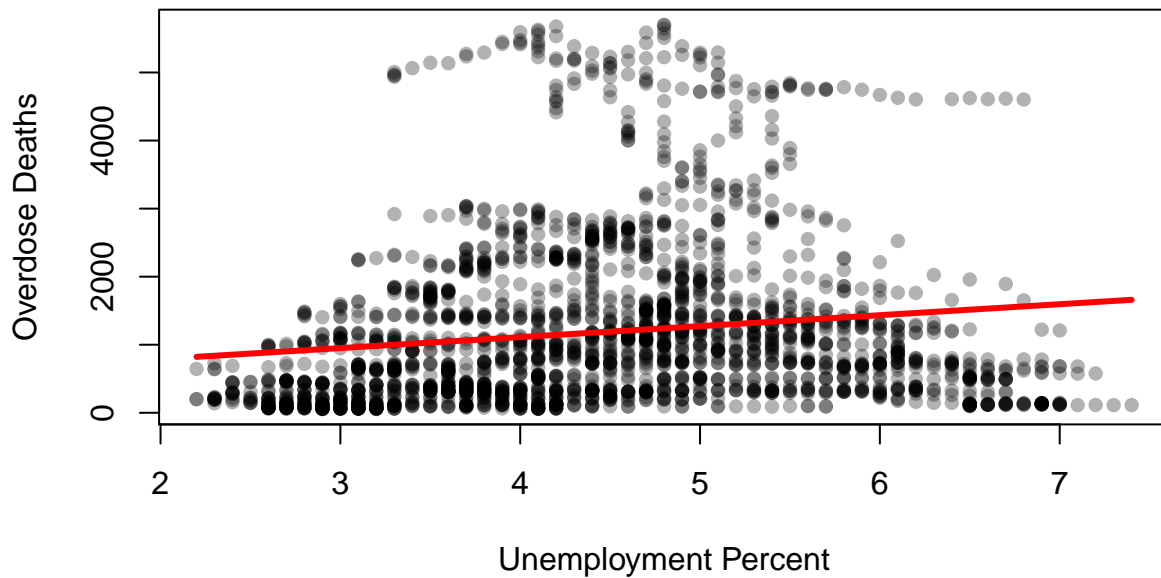
Baseline Model

```
##
## Call:
## lm(formula = overdoseDeaths ~ unemployment, data = overdose)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1546.8  -809.7  -360.1   351.2  4531.8
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    467.41     100.45   4.653 3.44e-06 ***
## unemployment    161.14      22.04   7.311 3.57e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1192 on 2446 degrees of freedom
## Multiple R-squared:  0.02139,    Adjusted R-squared:  0.02099
## F-statistic: 53.45 on 1 and 2446 DF,  p-value: 3.572e-13
```

The simple regression model has a positive coefficient for unemployment (155.25). With a t -statistic of 7.267 (p -value < 0.0001), this coefficient is very significant. The model has a positive association between unemployment and overdose deaths.

Overdose Deaths vs. Unemployment



Results

Conclusions and Decisions