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

Evan Arnold and Caleb Ren

12/06/2019

Introduction and Motivation

Exploratory Data Anslysis 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 in on a monthly basis.

The data of import 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 acordance 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, oxymorphone; 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 observation 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.

```
## [1] "3473513fa1cde2a0539bd2449749c5a3a4fc45ce"
```

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 avaiable 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.
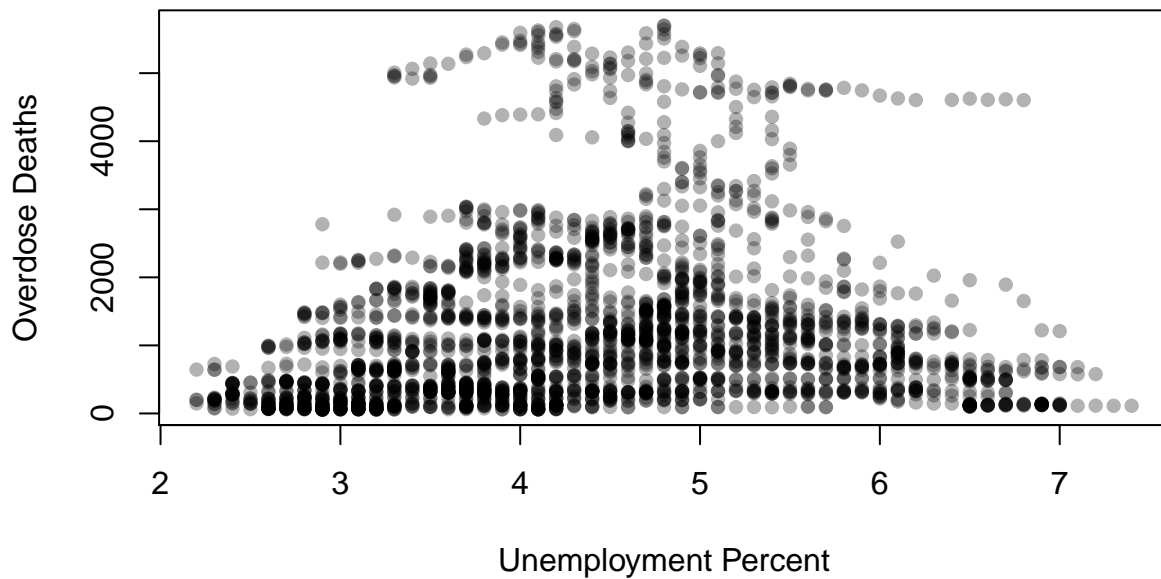
Using this definition, data was scraped from the BLS website and aggregated by each state and the District of Columbia. The unemployment rate in percent is given by the `unemployment` column. The lowest unemployment rate in a given state and month is Vermont in 2019 with a 2.1% unemployment rate. The 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%.

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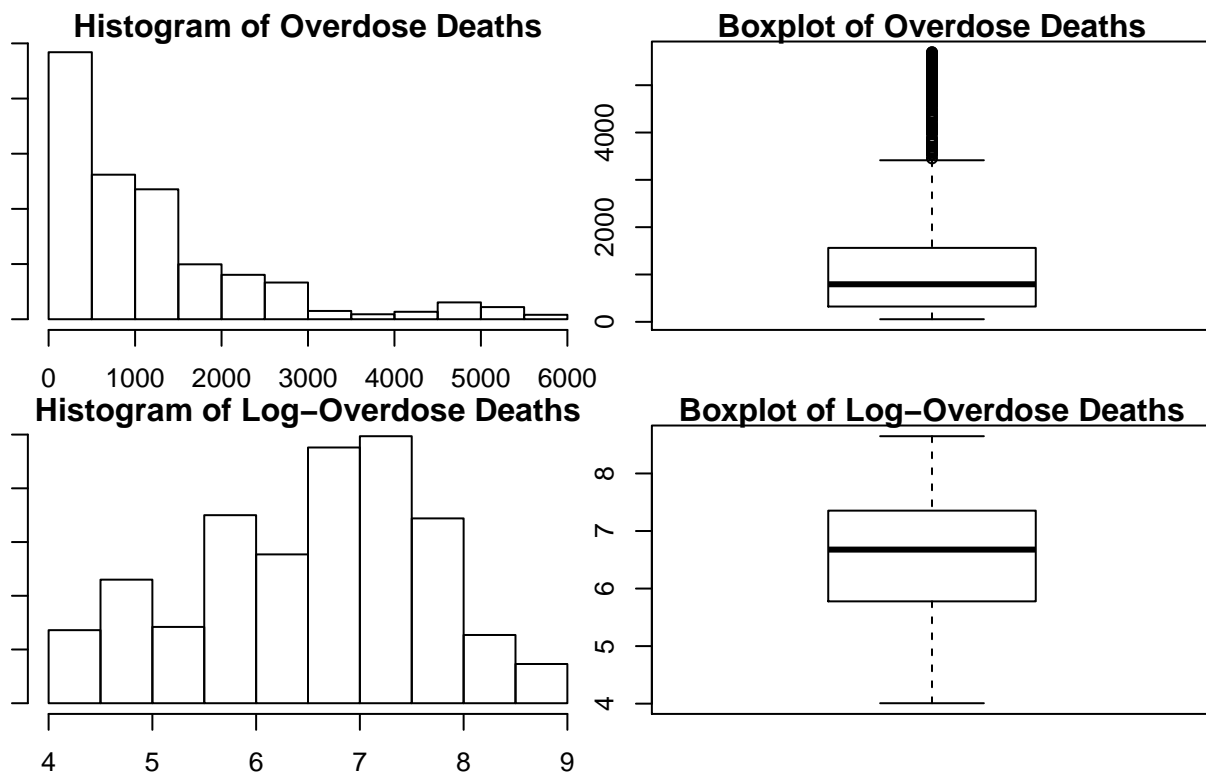
We see that the data is much closer to a Normal distribution if we apply a log transformation.


Baseline Model

```
##
```

```
## Call:
## lm(formula = overdoseDeaths ~ unemployment, data = overdose)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1545.9  -814.3  -374.3   366.3  4513.9
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    510.00      96.31   5.295 1.28e-07 ***
## unemployment   155.25      21.36   7.267 4.81e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1200 on 2650 degrees of freedom
## Multiple R-squared:  0.01954,    Adjusted R-squared:  0.01917
## F-statistic: 52.81 on 1 and 2650 DF,  p-value: 4.807e-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
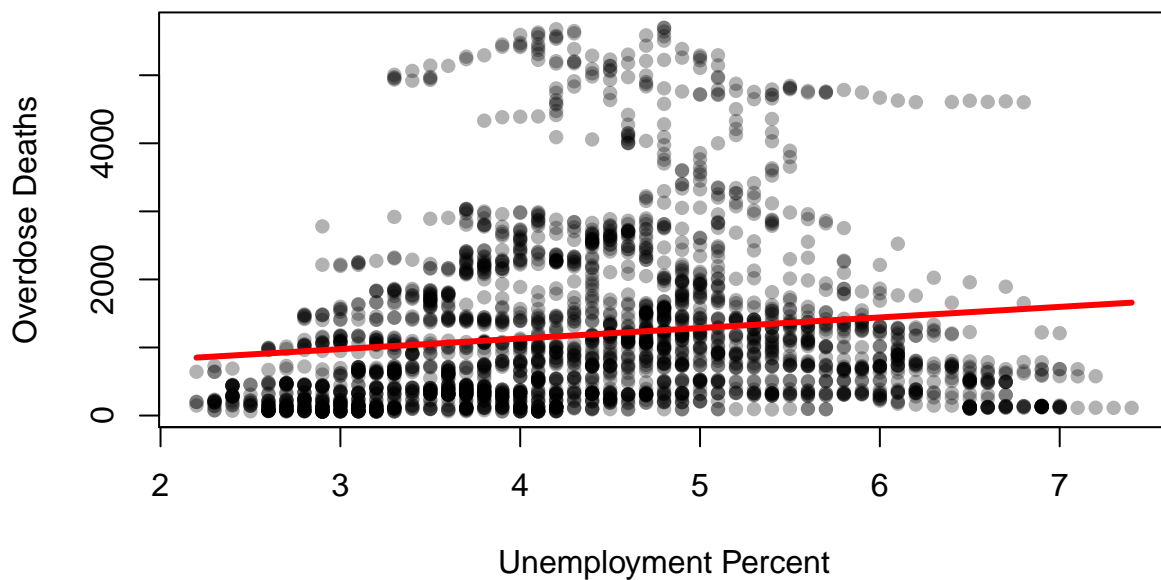
Results

Conclusions and Decisions