

Public Health Opioid Study

Mridul Khanna

Contents

Introduction	2
Theory	2
Data and Methodology	2
Data	2
Methodology	3
Data Cleaning and Transformation	3
Model Development	3
Exploratory Data Analysis	5
Scatter Plot: Standardised Poverty vs Mortality	5
Correlation between Poverty and Mortality	6
Scatter Plot: Opiates vs Mortality	6
Correlation between Opiates and Mortality	7
Histogram of Mortality Rate (Descriptive Distribution)	7
Regression Model	12
Justification for Linear Regression	13
Residual Plot	13
QQ Plot for the residuals	14
Regression model with Interaction	14
Does the effect of opiates on mortality get stronger in high-poverty counties?	14
Compare the Two Models using AIC	15
Mapping Regression Results	16
Complementary Model: Decision Tree Classification	18
Results and Discussion	19
Limitations	20
Conclusion	21
References	22

Introduction

Mortality caused by overdose (over dependence) of painkiller drugs such as opiate and opioid in the US has been an enormous societal and public health problem in the US since the mid-1990s. It has now been recognised as a “drug epidemic”. We will investigate the relationship between mortality and opiate drug prescription, and the potential cause of this drug epidemic.

Theory

McKelvie found that urbanized and impoverished neighbourhoods in North American cities have been some of the most impacted by drug use and the risk of overdose [1]. Lankenau et al found that in LA and NY, about half of the drug overdose or mortality population did not complete high school or were expelled from school, three quarters had a history of a psychological diagnosis, nearly all were homeless at some point or currently homeless [2].

Several papers concluded that one major cause of drug epidemic is the incredibly high number of opioid prescriptions since 1999 [3-8]. Reyes reported a case where when family physician refused to prescribe more painkiller medication, a patient found relief at pain management clinics and paid cash for up to 120 opiate-based pills during each visit [9]. Griesler et al found that prescription opioid misusers who misused both their own prescriptions and prescription opioid drugs not prescribed to them may be most at risk for overdose [10]. Dallin et al highlighted that the greatest challenge to overcome was a lack of legal mandates and proper education for healthcare providers on best practices for addressing the epidemic [11]. Brondeel et al suggested that an important aspect of opioid abuse prevention was the vigilant prescribing of medications with abuse potential [12]. Light in his book wrote about how physicians were been given usage education of new drugs and too many incentives to prescribe the drugs by pharmaceutical companies [13]. In 2014, US Drug Enforcement Administration ruled to limiting the legal supply of opioid drugs through narrowed circumstances in which opioids could be prescribed and stopped automatic repeat prescriptions.

The above two theories are supported by more researchers. Paulozzi et al found that rates of opioid misuse and overdose death are highest among poor, rural populations, and those who have mental illness; they also suggested improving medical practice in prescribing those drugs [14]. Fink et al listed income inequality, unemployment and poverty rate as few of the socio-economic indicators that associates with drug overdose and deaths. They concluded that rates of drug overdose deaths are concentrated in counties with greater prescription opioid supply [15]. Johnson et al named financial problems, unemployment, physical disability, depression, less educated and reside in rural areas most common factors of drug overdose or mortality; majority of decedents were prescribed opioids for management of chronic pain and many exhibited behaviours indicative of prescribed medication misuse [16].

Therefore, social-economic disadvantage and better drug supply and medical practice seemed to be the two main factors cause the drug epidemic. Researchers found many drug users are alcohol drinkers, however no causal relationship between alcohol abuse and drug overdose or mortality. Poverty is one element of social-economic disadvantage. Many who were prescribed for chronic pain management later ended up over dependent on the drugs.

We hypothesize that poverty and opioid prescription rates are positively associated with mortality in the United States.

Data and Methodology

Data

To prove the hypothesis, we have used county data from multiple sources in the U.S :

- Mortality rates from drug and alcohol use from the Centers for Disease Control and Prevention (CDC)
- Opiate prescription rates from national prescribing datasets
- Poverty rates from the U.S. Census Bureau American Community Survey (ACS)
- Geospatial shapefiles from the U.S. Census Bureau for mapping

These datasets were merged using a common county-level Id2/FIPS codes.

The variables used for the study are:

- Mortality rate: Deaths due to consumption of drugs and alcohol per 100,000 people (target variable)
- Poverty rate: It is the percentage of families living below the poverty level
- Opiate prescribing rate: Number of prescriptions per 100,000 people
- For the decision tree we are creating a binary poverty variable.

```
## [1] "Geography"
```

```
## [1] "County"
```

```
## [1] 1001 1003 1005 1007 1009 1011
```

```
## [1] 1001 1003 1005 1007 1009 1011
```

Methodology

Data Cleaning and Transformation

- Removing unreliable or missing mortality data
- Using z score standardisation of key variables
- Merged multiple datasets on FIPS/ID2 code

Model Development

Two modelling techniques were applied: 1. A multiple linear regression model was build. It was followed by including an interaction term between poverty and opiate prescribing rate. 2. A decision tree classification model using a binary version of poverty. This helped us to identify high-risk counties and segregating them into groups.

Both modelling methods were used to examine how poverty and access to opioid consumption is linked to mortality outcomes. We examined multiple visualisations and spatial data to produce maps that visualise risk levels and model predictions across U.S. counties.

```
##      Id2 below.poverty.rate      county mortality.rate
## 1 1001                9.4 Autauga, AL                12.5
## 2 1003                9.3 Baldwin, AL                22.6
## 3 1005               20.0 Barbour, AL                 9.0
## 4 1007               11.7  Bibb, AL                 14.1
## 5 1009               12.2 Blount, AL                18.1
## 6 1011               25.3 Bullock, AL                11.1
```

##	[1]	"10.0"	"10.1"	"10.2"	"10.3"	"10.4"
##	[6]	"10.5"	"10.6"	"10.7"	"10.8"	"10.9"
##	[11]	"11.0"	"11.1"	"11.2"	"11.3"	"11.4"
##	[16]	"11.5"	"11.6"	"11.7"	"11.8"	"11.9"
##	[21]	"113.4"	"12.0"	"12.1"	"12.2"	"12.3"
##	[26]	"12.4"	"12.5"	"12.6"	"12.7"	"12.8"
##	[31]	"12.9"	"13.0"	"13.1"	"13.2"	"13.3"
##	[36]	"13.4"	"13.5"	"13.6"	"13.7"	"13.8"
##	[41]	"13.9"	"14.0"	"14.1"	"14.2"	"14.3"
##	[46]	"14.4"	"14.5"	"14.6"	"14.7"	"14.8"
##	[51]	"14.9"	"15.0"	"15.1"	"15.2"	"15.3"
##	[56]	"15.4"	"15.5"	"15.6"	"15.7"	"15.8"
##	[61]	"15.9"	"16.0"	"16.1"	"16.2"	"16.3"
##	[66]	"16.4"	"16.5"	"16.6"	"16.7"	"16.8"
##	[71]	"16.9"	"17.0"	"17.1"	"17.2"	"17.3"
##	[76]	"17.4"	"17.5"	"17.6"	"17.7"	"17.8"
##	[81]	"17.9"	"18.0"	"18.1"	"18.2"	"18.3"
##	[86]	"18.4"	"18.5"	"18.6"	"18.7"	"18.8"
##	[91]	"18.9"	"19.0"	"19.1"	"19.2"	"19.3"
##	[96]	"19.4"	"19.5"	"19.6"	"19.7"	"19.8"
##	[101]	"19.9"	"20.0"	"20.1"	"20.2"	"20.3"
##	[106]	"20.4"	"20.5"	"20.6"	"20.7"	"20.8"
##	[111]	"20.9"	"21.0"	"21.1"	"21.2"	"21.3"
##	[116]	"21.4"	"21.5"	"21.6"	"21.7"	"21.8"
##	[121]	"21.9"	"22.0"	"22.1"	"22.2"	"22.3"
##	[126]	"22.4"	"22.5"	"22.6"	"22.7"	"22.8"
##	[131]	"22.9"	"23.0"	"23.1"	"23.2"	"23.3"
##	[136]	"23.4"	"23.5"	"23.6"	"23.7"	"23.8"
##	[141]	"23.9"	"24.0"	"24.1"	"24.2"	"24.3"
##	[146]	"24.4"	"24.5"	"24.6"	"24.7"	"24.8"
##	[151]	"24.9"	"25.0"	"25.1"	"25.2"	"25.3"
##	[156]	"25.4"	"25.5"	"25.6"	"25.7"	"25.8"
##	[161]	"25.9"	"26.0"	"26.1"	"26.2"	"26.3"
##	[166]	"26.4"	"26.5"	"26.6"	"26.7"	"26.8"
##	[171]	"26.9"	"27.0"	"27.1"	"27.2"	"27.3"
##	[176]	"27.4"	"27.5"	"27.6"	"27.7"	"27.8"
##	[181]	"27.9"	"28.0"	"28.1"	"28.2"	"28.3"
##	[186]	"28.4"	"28.5"	"28.6"	"28.7"	"28.8"
##	[191]	"28.9"	"29.0"	"29.1"	"29.2"	"29.4"
##	[196]	"29.5"	"29.6"	"29.7"	"29.8"	"29.9"
##	[201]	"30.0"	"30.1"	"30.2"	"30.3"	"30.5"
##	[206]	"30.6"	"30.7"	"30.8"	"30.9"	"31.0"
##	[211]	"31.1"	"31.2"	"31.3"	"31.4"	"31.5"
##	[216]	"31.6"	"31.7"	"31.8"	"31.9"	"32.3"
##	[221]	"32.4"	"32.5"	"32.6"	"32.7"	"32.8"
##	[226]	"32.9"	"33.0"	"33.1"	"33.2"	"33.3"
##	[231]	"33.4"	"33.5"	"33.6"	"33.7"	"33.8"
##	[236]	"33.9"	"34.0"	"34.1"	"34.2"	"34.3"
##	[241]	"34.4"	"34.5"	"34.6"	"34.8"	"34.9"
##	[246]	"35.0"	"35.1"	"35.3"	"35.4"	"35.6"
##	[251]	"35.7"	"35.9"	"36.0"	"36.1"	"36.2"
##	[256]	"36.3"	"36.4"	"36.5"	"36.7"	"36.8"
##	[261]	"36.9"	"37.0"	"37.2"	"37.3"	"37.4"
##	[266]	"37.5"	"37.7"	"37.9"	"38.0"	"38.1"

```
## [271] "38.2"      "38.3"      "38.4"      "38.5"      "38.6"
## [276] "38.7"      "38.8"      "38.9"      "39.1"      "39.2"
## [281] "39.4"      "39.5"      "39.6"      "39.7"      "39.9"
## [286] "4.2"       "4.6"       "4.9"       "40.0"      "40.4"
## [291] "40.5"      "40.6"      "40.7"      "40.8"      "41.0"
## [296] "41.2"      "41.5"      "41.7"      "41.8"      "41.9"
## [301] "42.1"      "42.2"      "42.3"      "42.4"      "42.5"
## [306] "42.9"      "43.2"      "43.3"      "43.5"      "43.6"
## [311] "43.9"      "44.1"      "44.2"      "44.5"      "44.6"
## [316] "44.7"      "44.9"      "45.0"      "45.1"      "45.4"
## [321] "45.5"      "45.7"      "45.8"      "45.9"      "46.2"
## [326] "46.7"      "46.9"      "47.0"      "47.5"      "47.9"
## [331] "48.2"      "48.5"      "49.0"      "49.1"      "49.7"
## [336] "49.9"      "5.3"       "5.8"       "5.9"       "50.0"
## [341] "50.1"      "50.2"      "50.3"      "51.0"      "51.4"
## [346] "51.5"      "51.8"      "51.9"      "52.2"      "52.7"
## [351] "52.9"      "53.0"      "53.1"      "53.4"      "53.8"
## [356] "54.5"      "54.6"      "55.0"      "55.7"      "55.8"
## [361] "56.0"      "58.4"      "59.2"      "6.2"       "6.3"
## [366] "6.5"       "6.7"       "6.8"       "60.3"      "61.6"
## [371] "61.8"      "64.3"      "66.8"      "67.2"      "67.3"
## [376] "67.4"      "68.9"      "7.1"       "7.2"       "7.3"
## [381] "7.4"       "7.5"       "7.6"       "7.7"       "7.8"
## [386] "7.9"       "70.1"      "74.5"      "78.9"      "8.0"
## [391] "8.1"       "8.2"       "8.3"       "8.4"       "8.5"
## [396] "8.6"       "8.7"       "8.8"       "8.9"       "80.2"
## [401] "83.7"      "86.3"      "87.1"      "9.0"       "9.1"
## [406] "9.2"       "9.3"       "9.4"       "9.5"       "9.6"
## [411] "9.7"       "9.8"       "9.9"       "90.8"      "97.6"
## [416] "Unreliable"
```

```
## Reading layer `cb_2016_us_county_20m' from data source
##   C:\Users\mridu\Desktop\Github\Data\US-data\US shapefiles\cb_2016_us_county_20m.shp'
##   using driver `ESRI Shapefile'
## Simple feature collection with 3220 features and 9 fields
## Geometry type: MULTIPOLYGON
## Dimension:      XY
## Bounding box:   xmin: -179.1743 ymin: 17.91377 xmax: 179.7739 ymax: 71.35256
## Geodetic CRS:   NAD83
```

```
## Joining with `by = join_by(county)`
```

```
## Warning in sf_column %in% names(g): Detected an unexpected many-to-many relationship between `x` and
## i Row 1205 of `x` matches multiple rows in `y`.
## i Row 1138 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

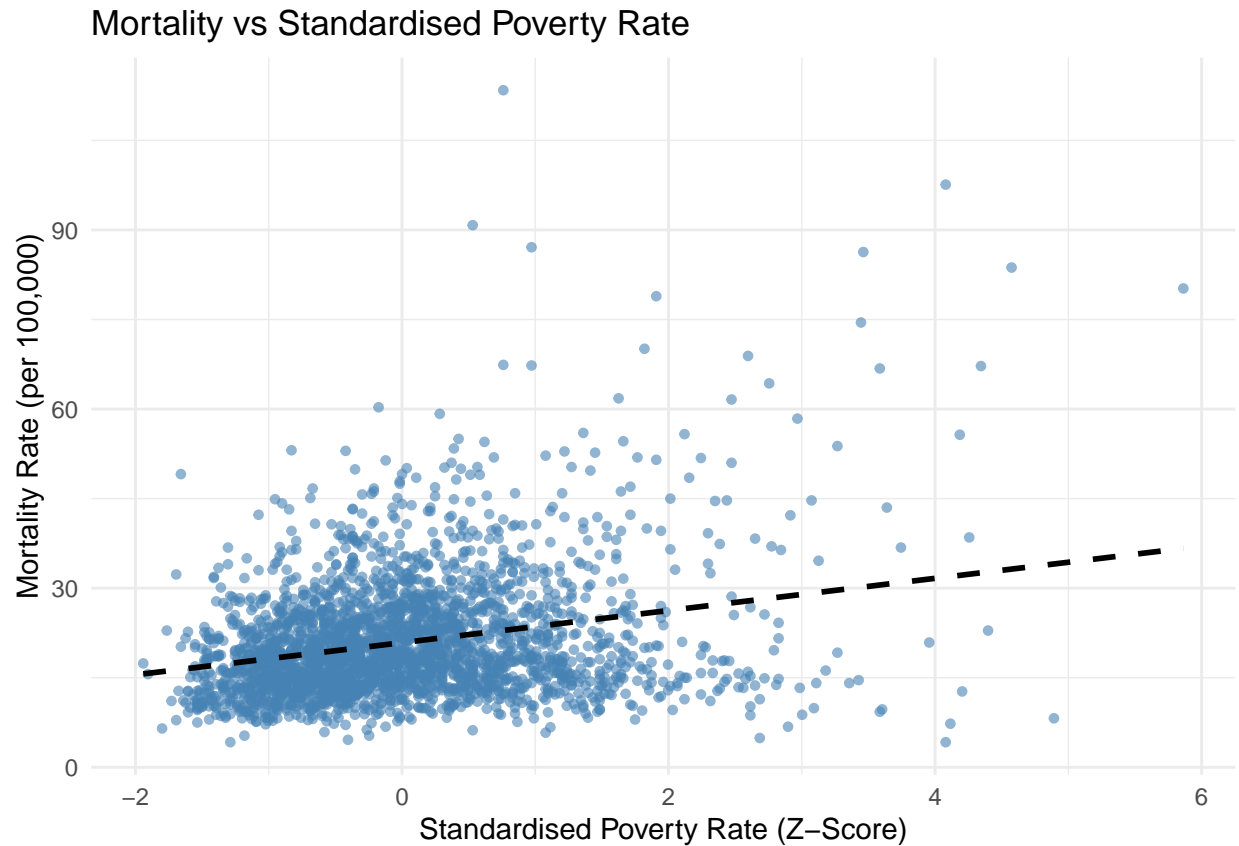
Exploratory Data Analysis

Scatter Plot: Standardised Poverty vs Mortality

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 275 rows containing non-finite outside the scale range
## (`stat_smooth()`).

## Warning: Removed 275 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



The above graph shows positive correlation. As poverty increases, mortality rate also shows tendency to increase. The plot shows a lot of values are clustered around the average value of poverty rate.

Certain outliers are evident, with a limited number of counties showing either exceedingly high poverty levels ($Z > 4$) or exceptionally high death rates ($Y > 80$). Since poverty appears to be a moderate cause of death, it should be taken into account when developing the model.

Correlation between Poverty and Mortality

```
## [1] 0.2641112
```

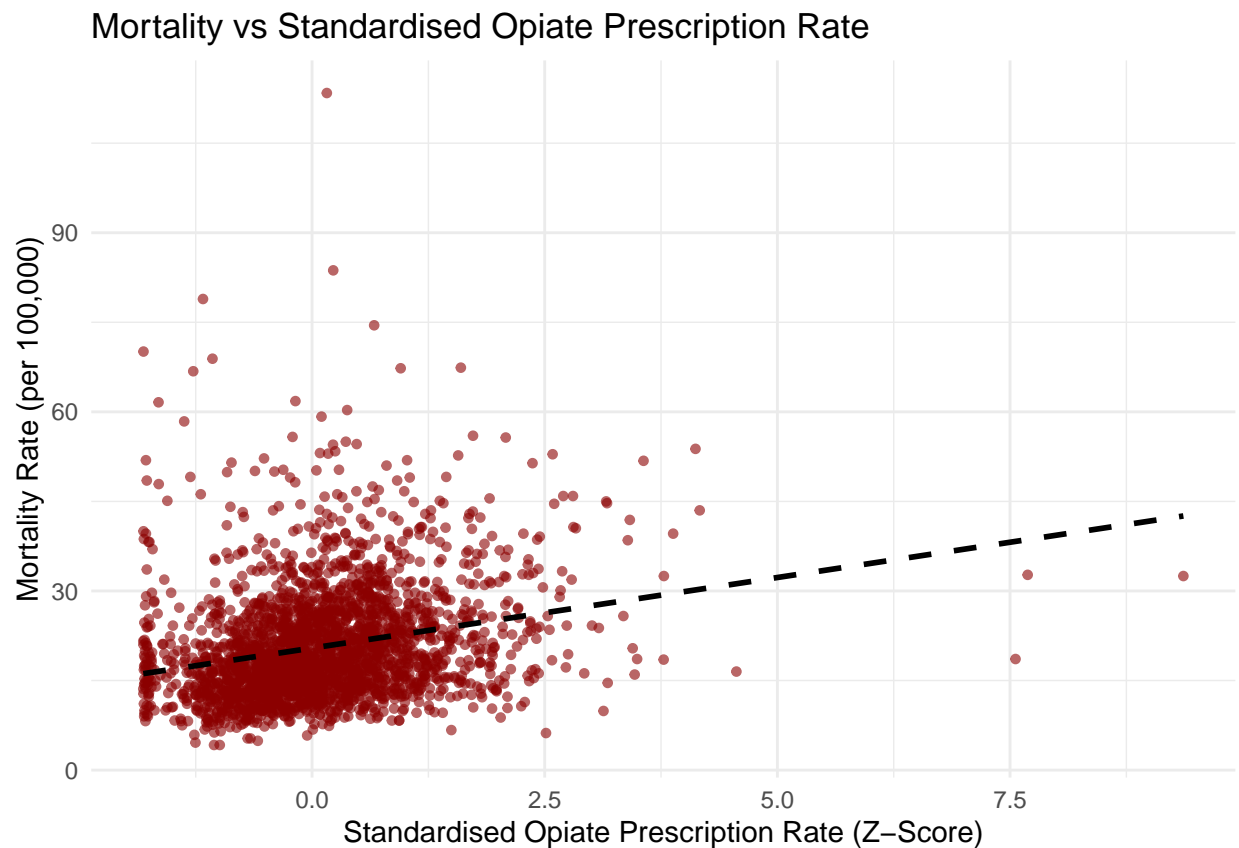
The correlation is 0.26 which indicates a positive weak correlation between poverty and mortality.

Scatter Plot: Opiates vs Mortality

```
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 316 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

```
## Warning: Removed 316 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



The above graph indicates a positive weak correlation between opiate prescribing and drug-related mortality. The plot shows most counties have prescription rate around the mean value or slightly higher.

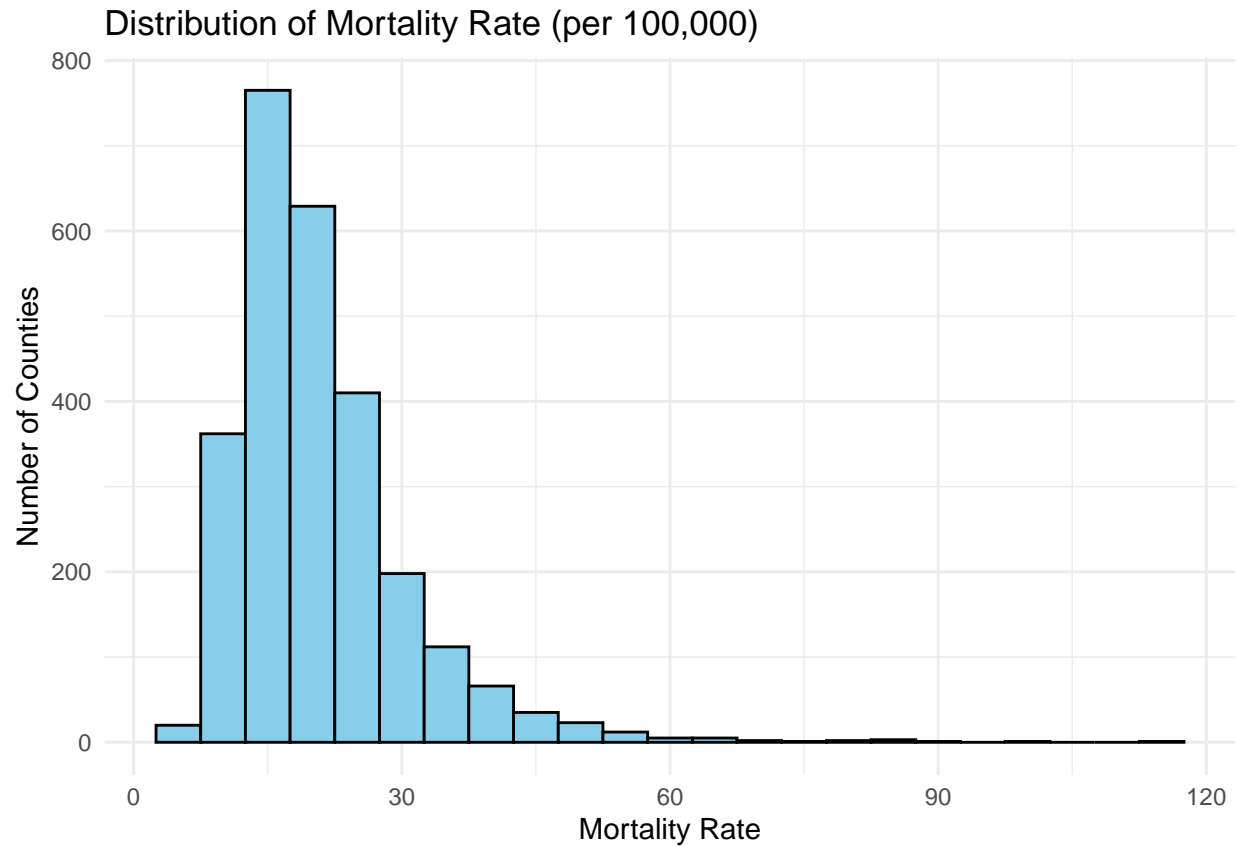
There are some outliers present, indicating very high prescription rates, but not all of them have high mortality rate. It shows moderate effect and should be included in model development, but suggest presence of other factors as well.

Correlation between Opiates and Mortality

```
## [1] 0.2437066
```

Histogram of Mortality Rate (Descriptive Distribution)

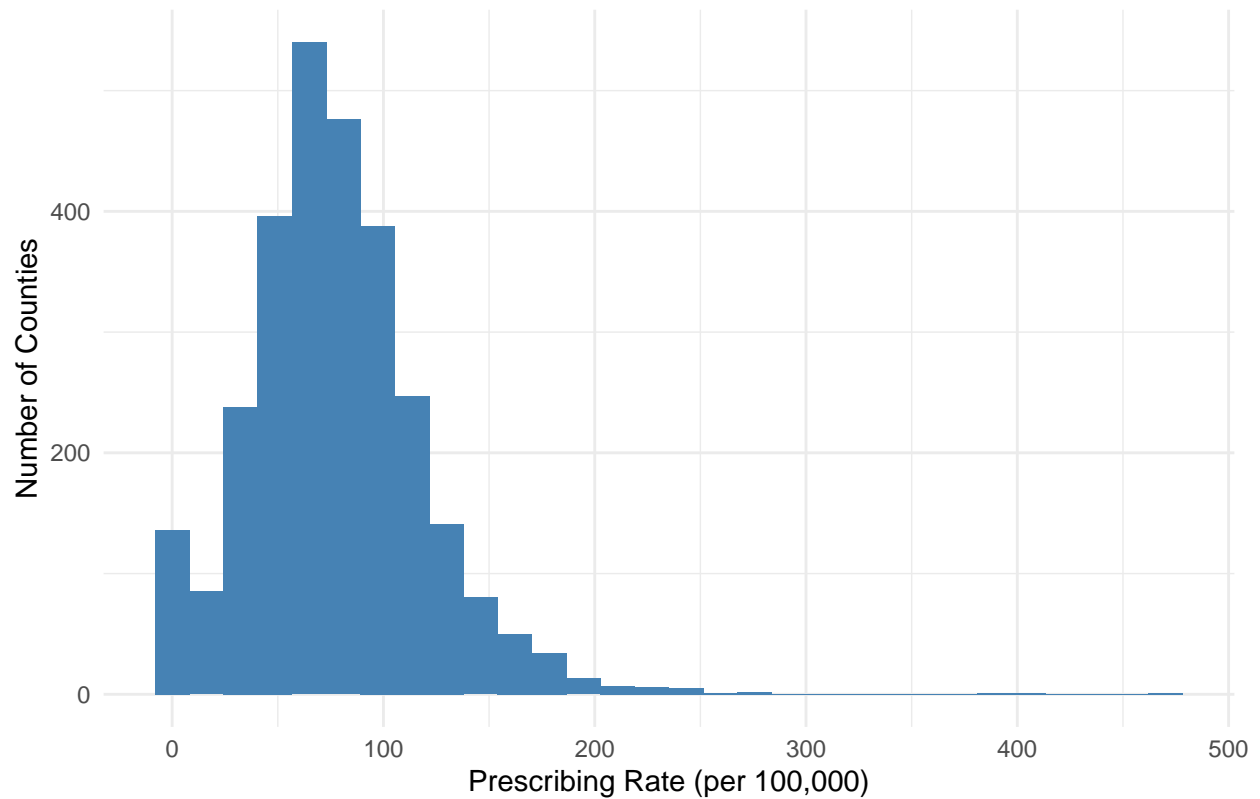
```
## Warning: Removed 275 rows containing non-finite outside the scale range
## (`stat_bin()`).
```



The plot is right skewed. Most values are clustered between 15 and 40. The warnings shows 275 rows removed, as we had converted “Unreliable” to NA. Some counties have higher rates of mortality, as depicted by the skewness.

```
## Warning: Removed 80 rows containing non-finite outside the scale range
## (`stat_bin()`).
```


Distribution of Opioid Prescribing Rates

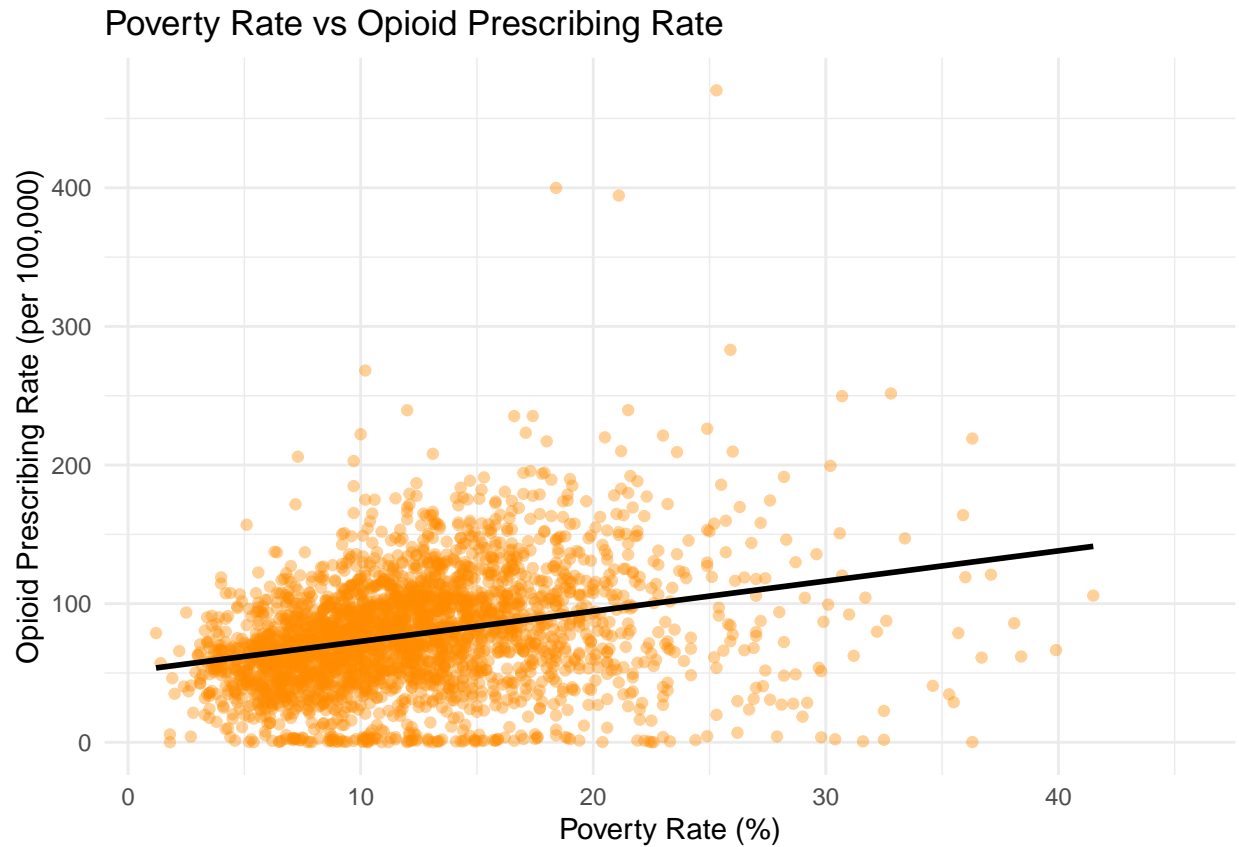


Most counties have prescribing rates between 50 and 150 per 100,000 people, with a few extreme outliers exhibiting rates over 400.

```
## `geom_smooth()` using formula = 'y ~ x'
```

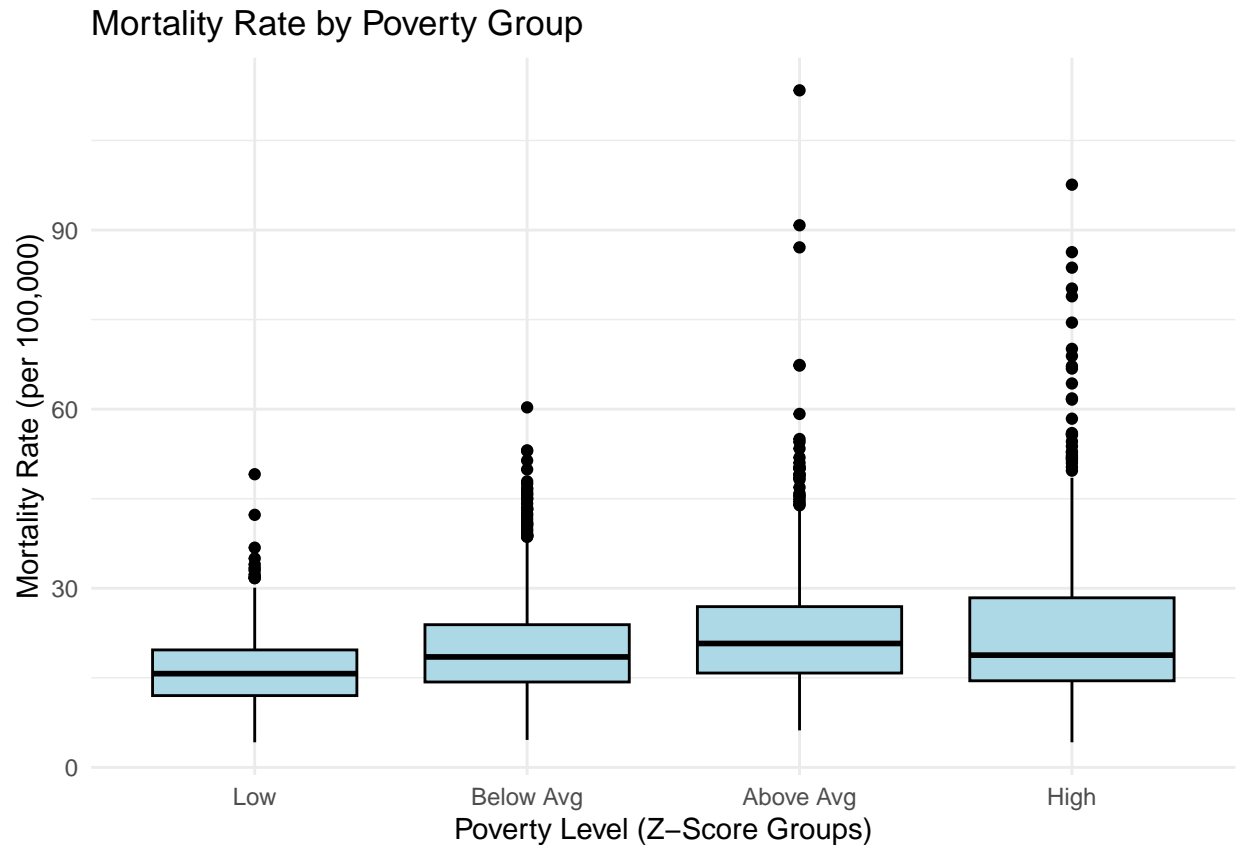
```
## Warning: Removed 80 rows containing non-finite outside the scale range  
## (`stat_smooth()`).
```

```
## Warning: Removed 80 rows containing missing values or values outside the scale range  
## (`geom_point()`).
```



The scatter plot demonstrates a weak to moderate positive relationship between poverty and opioid prescribing rates across counties in the U.S. The Pearson correlation coefficient is 0.291, indicating that higher poverty levels are modestly associated with higher rates of opioid prescriptions.

```
## Warning: Removed 275 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



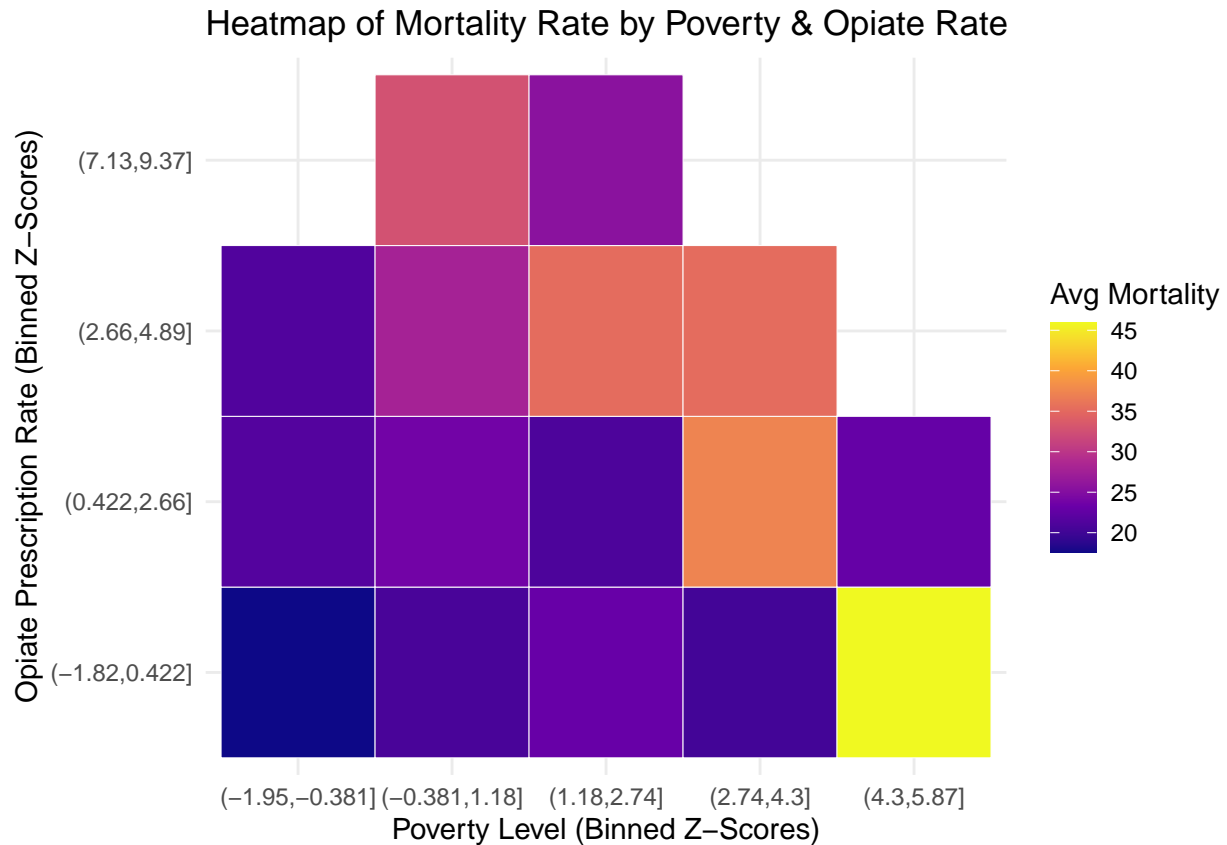
The above plot shows that poverty rate increases -> average mortality rate increases. It shows relationship between mortality rate and poverty.

The number of outliers for above average or high poverty level is extremely high. It could be due to the influence of confounding factors.

Poverty should be included in the development of our model.

Heatmap: Mortality by Standardised Poverty and Opiate Rates

```
## `summarise()` has grouped output by 'poverty.bin'. You can override using the
## `.groups` argument.
```



The above heatmap shows that mortality increases as either poverty or opiate prescribing increases — but the highest mortality occurs when both are high. This shows poverty and opiates have a combined impact on mortality, not just separate effects.

Regression Model

We are building a regression model to understand the linear relation between poverty, opiate prescription rate and mortality across counties due to consumption of alcohol and drugs.

```
##
## Call:
## lm(formula = mortality.rate ~ z.poverty + z.opiates, data = county.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.983  -5.824  -1.562   3.963  91.399
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  20.4864    0.1757  116.591  <2e-16 ***
## z.poverty     1.6040    0.1941   8.265  <2e-16 ***
## z.opiates     1.8630    0.1916   9.724  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 8.934 on 2609 degrees of freedom
## (316 observations deleted due to missingness)
## Multiple R-squared:  0.08339,    Adjusted R-squared:  0.08269
## F-statistic: 118.7 on 2 and 2609 DF,  p-value: < 2.2e-16
```

The above summary shows:

The co-efficients for both the variables are positive. This shows that when poverty and prescription rates are increasing, mortality rate is also increasing.

The p value (< 0.05) of both the variables suggests that they are contributing significantly to mortality.

R-squared = 0.0834 \rightarrow about 8.3% of variation in mortality is explained by poverty + opiates. This shows that drug mortality is affected by various factors, but the higher poverty and more prescriptions are increasing the mortality rate as found out in the descriptive analysis.

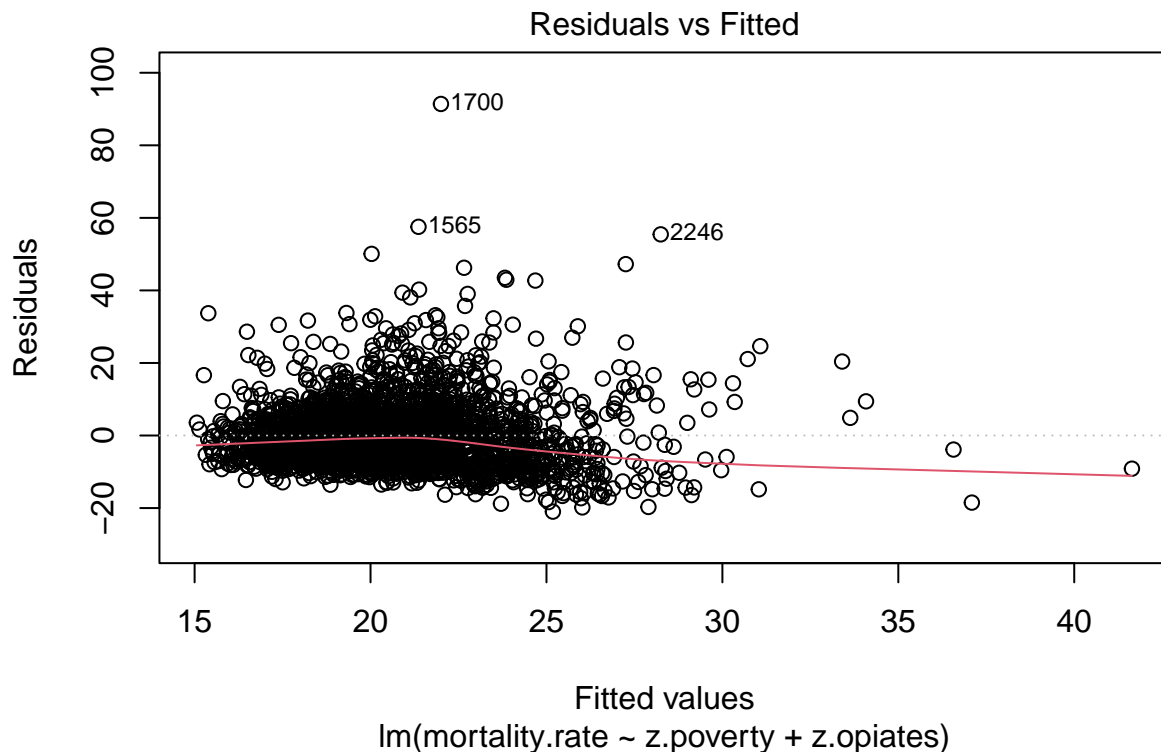
This is a reasonable value for social science data.

Justification for Linear Regression

We selected linear regression because our dependent variable mortality rate is a continuous numeric variable. Linear regression is well-suited for modeling continuous outcomes and allows us to estimate the marginal effect of each predictor (e.g., poverty rate, prescribing rate) on mortality.

To check if linear regression is a valid choice for this data:

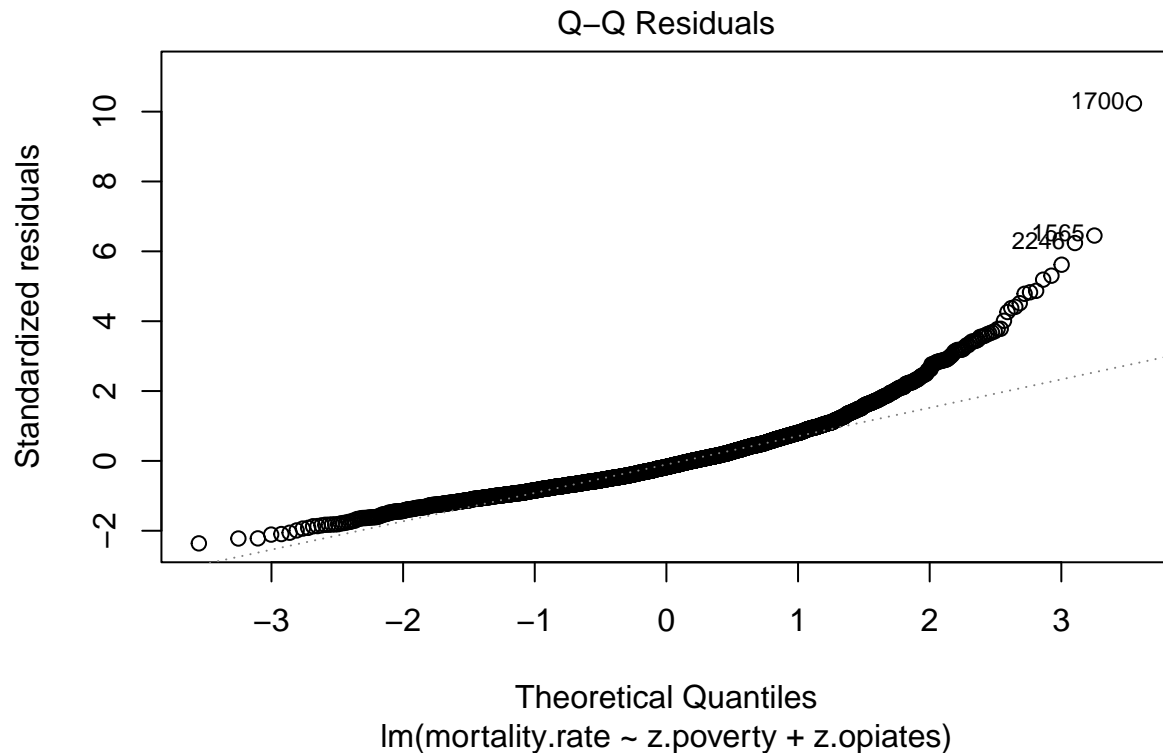
Residual Plot



The above plot shows that linear regression is working for this data. Most of the errors are spread around 0, and are evenly spread indicating the model is predicting well.

Some points are far away, indicating for some counties the model didn't perform well, which is expected in the real world.

QQ Plot for the residuals



The above plot is checking if the errors i.e. residuals of the model are normally distributed, which is expected for linear regression models. If it is valid, then the residual points (black dots) would follow the dotted line.

As seen from the graph, most of the points follow normal distribution as expected, with a few large positive outliers towards the right showing model unpredicted by large value (curve bending up). This could be due to some counties having higher than expected mortality. As in real-world data, some outliers are expected, especially with social science data.

The assumption of residuals following normal distribution is reasonable and linear regression is valid for this data.

The above discussions and results show that linear regression is appropriate for our data.

Regression model with Interaction

Does the effect of opiates on mortality get stronger in high-poverty counties?

Here we are testing whether the relationship between opiates prescriptions and mortality depends on county-level poverty. i.e. interaction : $\text{z.poverty}:\text{z.opiates}$

```
##
## Call:
## lm(formula = mortality.rate ~ z.poverty * z.opiates, data = county.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23.360  -5.771  -1.572   3.875  91.266
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    20.6362     0.1793 115.083 < 2e-16 ***
## z.poverty       1.6100     0.1935   8.319 < 2e-16 ***
## z.opiates       2.2005     0.2094  10.507 < 2e-16 ***
## z.poverty:z.opiates -0.6272     0.1594  -3.934 8.58e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.909 on 2608 degrees of freedom
## (316 observations deleted due to missingness)
## Multiple R-squared:  0.0888, Adjusted R-squared:  0.08775
## F-statistic: 84.72 on 3 and 2608 DF,  p-value: < 2.2e-16
```

The coefficient for the interaction term i.e $z.poverty:z.opiates = -0.6272$ is negative, which implies that the effect of opiate prescription is weaker with increasing poverty.

The p value for the interaction term i.e. $8.58e-05$ shows it is statistically significant as $p < 0.001$.

The impact of opiate prescriptions depends on the level of poverty

Low poverty counties: It is observed that for these counties opiate prescription is linked strongly to high mortality rates due to drugs and alcohol.

High poverty counties: It is observed that for these counties the opiate prescription rate is linked weakly to high mortality rates due to drugs and alcohol.

To conclude, both opiate prescription and poverty are increasing mortality. In a case when both of them are high, the combined effect is not a simple linear stack up. Say for a county with high poverty, the risk imposed by high opiate prescription becomes lesser.

Compare the Two Models using AIC

To show that the interaction improves model performance:

```
##              df      AIC
## reg.model      4 18857.39
## reg.model.int  5 18843.93
```

The AIC metric shows how well our model is fitting the data, by penalizing the complexity factor. A lower AIC Value for the interaction portrays better model, even though it is slightly more complex.

To conclude:

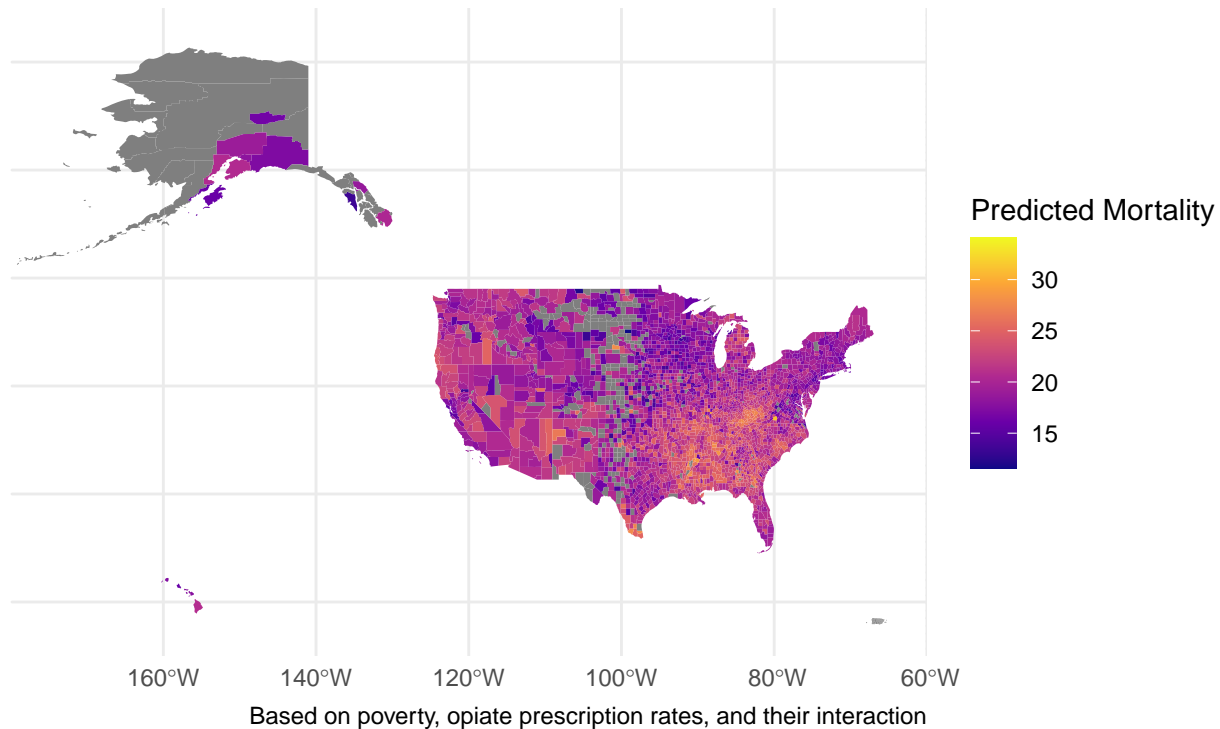
The combined effect of poverty and opiates explains mortality more accurately.

Mapping Regression Results

Here we are mapping the predicted mortality rates for different US Counties.

```
## Joining with `by = join_by(county, below.poverty.rate, mortality.rate,  
## X2016.Prescribing.Rate)`
```

Predicted Mortality Rate by County (Model-Based)



Based on our final model including the interaction, we have plotted the above graph showing mortality rates. The predictions from the model were mapped with the US spatial data.

Purple = lower deaths ; Yellow = higher deaths

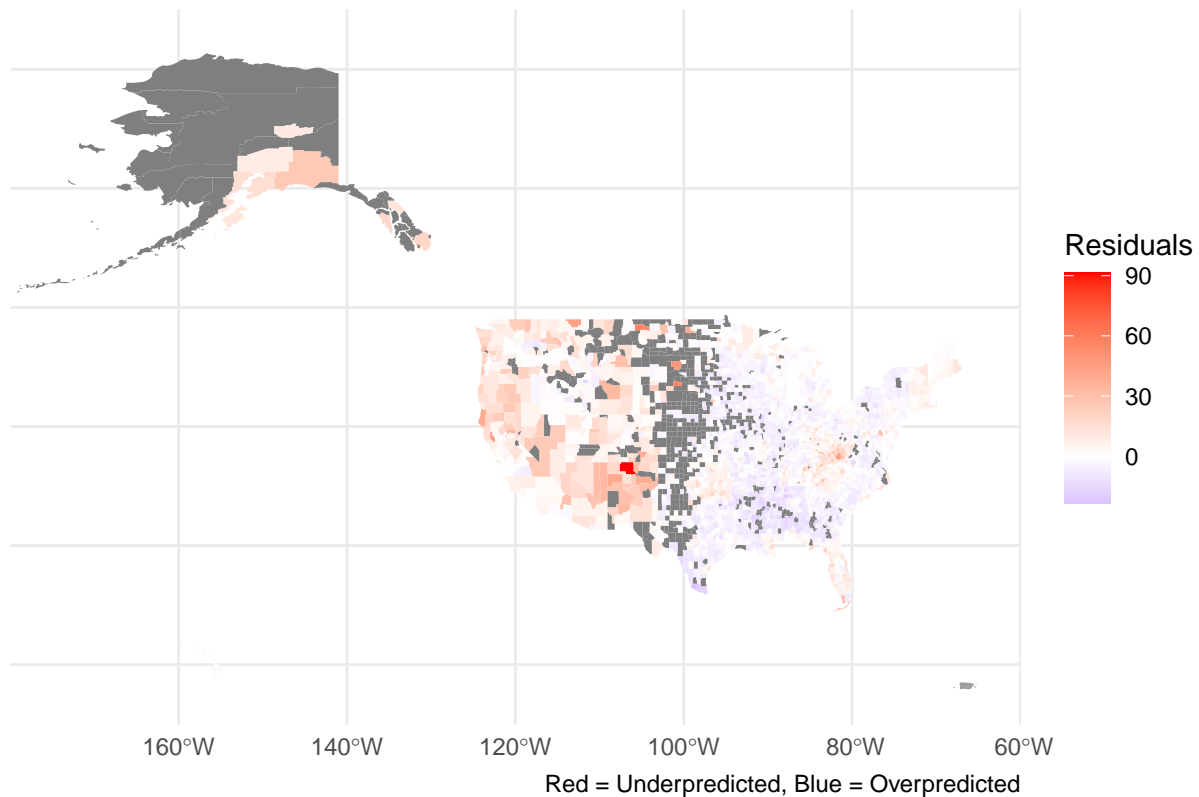
The above map shows that counties in the Alpine, Midwest and Southeast have high mortality. These are often places that have high poverty and higher opiate prescription rates.

This map also shows us which areas are most at risk, even if some counties had missing data, which makes it helpful for policy making.

```
## Joining with `by = join_by(county)`
```

```
## Warning in sf_column %in% names(g): Detected an unexpected many-to-many relationship between `x` and  
## i Row 1205 of `x` matches multiple rows in `y`.  
## i Row 1138 of `y` matches multiple rows in `x`.  
## i If a many-to-many relationship is expected, set `relationship =  
## "many-to-many"` to silence this warning.
```

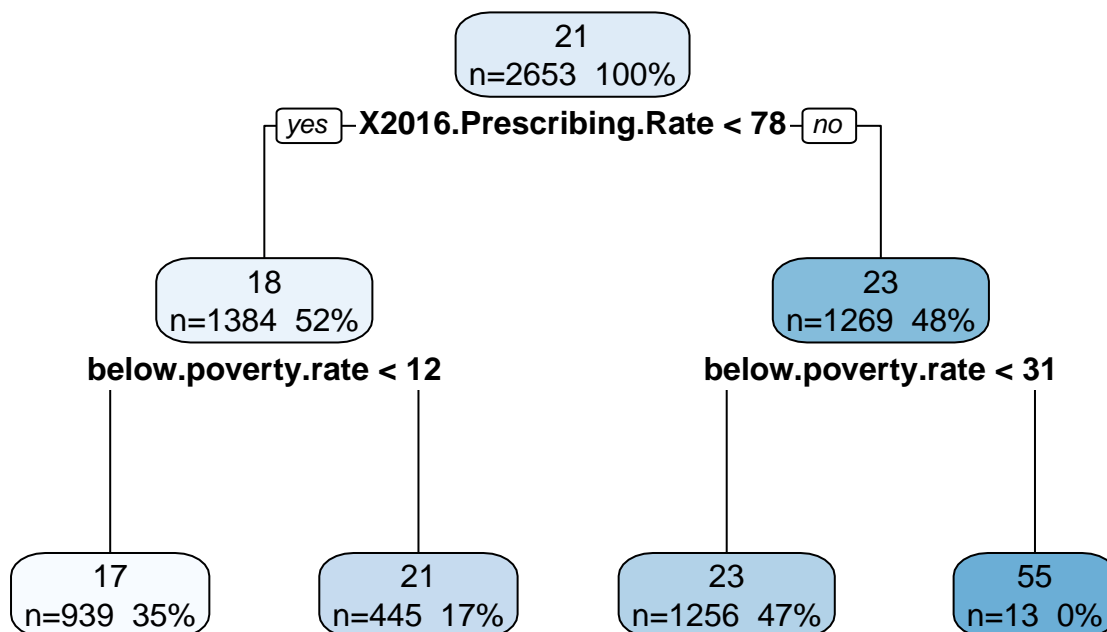

Model Residuals by County



In the above plot: 1) Red areas show where our model underpredicted → Actual deaths were higher than predicted. This is evident in some Midwestern and Western countries. This could be due to other factors imposing risk apart from poverty and opiate prescription rate.

- 2) Blue areas show where our model overpredicted → Actual deaths were lower than predicted. This is evident in southern or eastern countries. This could be due to good healthcare services or healthy lifestyle awareness.
- 3) And white/grey areas show a value close to zero → Model predicted pretty well. This is evident in most areas showing the model did a good job.

Complementary Model: Decision Tree Classification

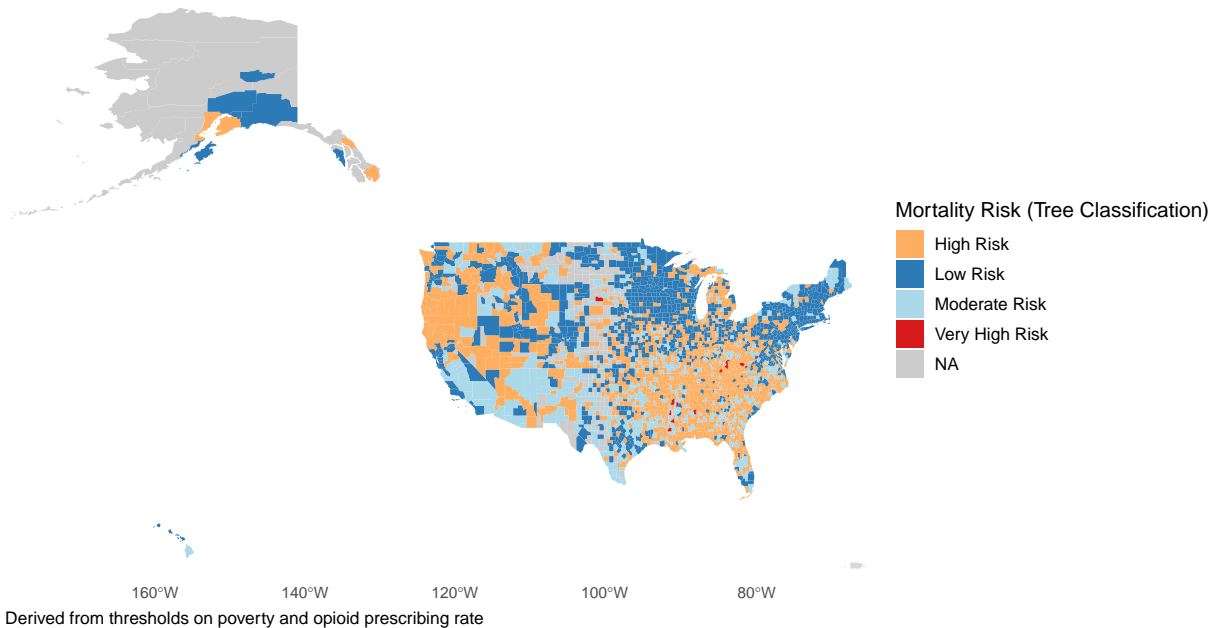


The regression tree reveals interpretable thresholds for both opioid prescribing rates and poverty levels. Counties with low prescribing rates (< 78) and low poverty (< 12%) tend to have the lowest mortality rates (~17 per 100,000). Conversely, counties with high prescribing rates (≥ 78) and very high poverty (≥ 31%) have dramatically higher mortality rates, with a subgroup averaging 55 deaths per 100,000. This supports the hypothesis that opioid access and socioeconomic disadvantage interact to influence mortality outcomes, and that the risk increases disproportionately once certain thresholds are crossed.

```
## Warning in sf_column %in% names(g): Detected an unexpected many-to-many relationship between `x` and
## i Row 1205 of `x` matches multiple rows in `y`.
## i Row 1138 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```
## Warning in sf_column %in% names(g): Detected an unexpected many-to-many relationship between `x` and
## i Row 1205 of `x` matches multiple rows in `y`.
## i Row 1138 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

County Classification Based on Decision Tree Rules



The map illustrates a decision-tree-based classification of U.S. counties into four mortality risk levels using poverty rate and opioid prescribing rate as predictors. Several key spatial patterns emerge: Low Risk counties (blue) are predominantly located in the Northeast and parts of the Midwest, regions typically characterized by lower poverty rates and relatively controlled opioid prescribing. This aligns with expectations and supports the model's validity. High Risk counties (orange) are widespread across the South, Midwest, and Western states, indicating a broader regional prevalence of higher opioid prescribing or moderate poverty. These areas may represent underserved or medically vulnerable populations. Very High Risk counties (red) are relatively rare but form visible clusters in the South and Appalachia, suggesting that combined socioeconomic disadvantage and high opioid access significantly elevate mortality risk. These regions may warrant targeted policy attention or public health intervention. Moderate Risk counties (light blue) fill in the transitional spaces—places where poverty is high but prescribing is relatively low, indicating that structural disadvantage alone, even without excessive opioid access, still correlates with elevated risk. Gray counties (NA) likely reflect missing or unreliable data. They cluster in Alaska, parts of the West, and scattered counties across the South, underscoring the need for improved data coverage in certain regions.

Results and Discussion

We started by building a basic linear regression model using poverty rate and opioid prescribing rate as the two main predictors. The results showed that both variables had a significant positive relationship with mortality, meaning that counties with higher poverty and more opioid prescriptions tended to have higher drug and alcohol related death rates, even after accounting for other factors.

The model had an adjusted R^2 of about 0.083, which means it explained around 8% of the variation in mortality. While this is relatively low, it's still acceptable for social science research and suggests that other unmeasured factors also play a role.

Based on the fitted values from this model, we created a county-level map showing predicted mortality rates. Darker colors on the map indicate higher predicted mortality. The results highlighted that counties in the southeastern and south-central parts of the U.S., especially those with higher poverty, were more likely to have elevated predicted mortality. In contrast, counties in the northeast and coastal western regions tended to have lower predicted mortality. This map also helped fill in areas where actual data were missing or labeled as “Unreliable,” providing a more complete nationwide picture.

In the second model, we added an interaction term between standardized poverty and prescribing rates ($z.\text{poverty} * z.\text{opiates}$). The interaction was significant, indicating that in counties with higher poverty, the effect of opioid prescribing on mortality was even stronger. In other words, the combined presence of poverty and high prescription levels can make the health risks worse.

We used this second model to generate another prediction map. Compared to the first map, this one showed even higher predicted mortality in high-poverty, high-prescription areas—particularly in the inland southeastern U.S. This supports the idea that the model effectively captured the interaction between the two variables and reinforces the importance of considering overlapping structural risks when forming public health responses.

Together, both maps showed how the predicted mortality risk varies across space and helped fill in counties with missing or unreliable data. They also clearly revealed that economic disadvantage and medical risk—poverty and opioid access—often overlap geographically, especially in parts of the South and Midwest.

We also used decision trees to classify counties according to below poverty rate as a binary variable, and the results showed that when the poverty rate exceeded 31% and the prescription rate was high, the mortality rate increased significantly, averaging more than 50 deaths per 100,000 people. This classification showed obvious geographical clustering, further supporting the conclusion of the linear model.

Although the linear models assume a simple relationship and don’t account for spatial correlation, combining them with map visualizations gave us valuable insights into how social and medical risks overlap geographically. Future research could improve on this by adding more variables—like race, education, or healthcare access—or by using spatial regression or machine learning methods. From a policy perspective, our findings suggest that resources should be prioritized in areas of the Southeast and Midwest where poverty and high prescription rates intersect.

Limitations

Social-economic disadvantage includes many elements, our model only highlighted poverty. Other elements can be further investigated individually or combined (to understand the joint impact). For example, Fink et al concluded that income inequality, Rey index and HDI showed the most consistent and substantial associations with drug overdose deaths, whereas unemployment and poverty rates showed weaker associations [15].

Legal opiate prescription practice in different counties(states) varies, which might cause different causal correlation to mortality rate. It is impossible to capture illicit opiate usage, which would most definitely cause different result. Martin et al proved in July 2016 sales of opioids through US crypto markets represented 13.7% of all drug sales compared with a modelled estimate of 6.7% of all sales had the new schedule not been introduced [25].

One key limitation of the linear regression model is that it assumes a linear relationship between the predictors (poverty rate and opioid prescribing rate) and the outcome (mortality rate). However, in real-world data, especially in social and health context, these relationships are often non-linear or involve threshold effects. This means the linear model may underfit the data and fail to capture more complex interactions.

Moreover, binary classification of poverty discards nuance, although binary splitting improves interpretability, it removes the ability to detect non-linear or threshold effects within continuous variation in poverty rates.

Conclusion

We examined how **county-level poverty and opiate prescription rates** influence mortality. Through scatter plots, correlations, and regression modeling, we found consistent evidence of **positive associations** — higher poverty and prescription rates were both linked to elevated mortality. The inclusion of an **interaction term** revealed that these effects confound each other in disadvantaged regions.

We found that poverty and opiate prescriptions both increase drug-related deaths — and when both are high, it is especially bad. Our model predicts these deaths across the US and maps show some regions are hit much harder than others.

Our mapping of predicted values and residuals added spatial context, revealing clusters of high-risk areas and regions where the model underperformed — potentially due to missing social or healthcare-related variables.

These findings highlight the **structural nature of drug-related mortality**, emphasizing the need for place-based public health policies that tackle poverty and prescription practices in tandem. While the model performed well overall, further work could incorporate demographic, behavioral, or access-to-care variables to improve accuracy.

References

1. McKelvie, S. (2020). “Smack in the middle”: Urban governance and the spatialization of overdose epidemics. *City & Community*, 19(3), 704–725. <https://doi.org/10.1111/cico.12456>
2. Lankenau, S. E., Teti, M., Silva, K., Bloom, J. J., Harocopos, A., & Treese, M. (2012). Patterns of prescription drug misuse among young injection drug users. *Journal of Urban Health*, 89(6), 1004–1016. <https://doi.org/10.1007/s11524-012-9691-9>
3. Walde, J., Andersson, L., Johnson, B., & Håkansson, A. (2023). Drug prescriptions preceding opioid-related deaths—a register study in forensic autopsy patients. *PLOS ONE*, 18(5), e0285583. <https://doi.org/10.1371/journal.pone.0285583>
4. Rudd, R. A., Seth, P., David, F., & Scholl, L. (2016). Increases in drug and opioid overdose deaths—United States, 2000–2015. *Morbidity and Mortality Weekly Report*, 65, 1445–1452.
5. Florence, C. S., Zhou, C., Luo, F., & Xu, L. (2016). The economic burden of prescription opioid overdose, abuse, and dependence in the United States, 2013. *Medical Care*, 54(10), 901–906.
6. Hah, J. M., Bateman, B. T., Ratliff, J., Curtin, C., & Sun, E. (2017). Chronic opioid use after surgery: Implications for perioperative management in the face of the opioid epidemic. *Anesthesia & Analgesia*, 125(5), 1733–1740.
7. Vadivelu, N., Kai, A. M., Kodumudi, V., Sramcik, J., & Kaye, A. D. (2018). The opioid crisis: A comprehensive overview. *Current Pain and Headache Reports*, 22(3), 16.
8. Wunsch, H., Wijeyesundera, D. N., Passarella, M. A., & Neuman, M. D. (2016). Opioids prescribed after low-risk surgical procedures in the United States, 2004–2012. *JAMA*, 315(15), 1654–1657.
9. Reyes, R. (2010, October 31). Pain clinics feeding habit. *McClatchy - Tribune Business News*. <https://www.proquest.com/wire-feeds/pain-clinics-feeding-habit/docview/761207402/se-2?accountid=14757>
10. Griesler, P. C., Hu, M., Wall, M. M., & Kandel, D. B. (2019). Medical use and misuse of prescription opioids in the US adult population: 2016–2017. *American Journal of Public Health*, 109(9), 1258–1265. <https://doi.org/10.2105/AJPH.2019.305162>
11. Dallin, J., King, C. R., & Curtis, G. (2023). The opioid epidemic: A review of the contributing factors, negative consequences, and best practices. *Cureus*, 15(7). <https://doi.org/10.7759/cureus.41621>
12. Brondeel, K. C., Malone, K. T., Ditmars, F. R., Vories, B. A., Ahmadzadeh, S., Tirumala, S., ... Kaye, A. D. (2023). Algorithms to identify nonmedical opioid use. *Current Pain and Headache Reports*, 27(5), 81–88. <https://doi.org/10.1007/s11916-023-01104-7>
13. Light, D. (2010). *The risks of prescription drugs*. Columbia University Press. <https://doi.org/10.7312/ligh14692>
14. Paulozzi, L., Franklin, G., Kerlikowske, R. G., Jones, C. M., Ghiya, N., & Popovic, T. (2012). CDC grand rounds: Prescription drug overdoses—A U.S. epidemic. *Morbidity and Mortality Weekly Report*, 61(1), 10–13. <https://link.gale.com/apps/doc/A278631245/AONE?u=usyd&sid=bookmark-AONE&xid=cb6c6511>
15. Fink, D. S., Keyes, K. M., Branas, C., Cerdá, M., Gruenwald, P., & Hasin, D. (2023). Understanding the differential effect of local socio-economic conditions on the relation between prescription opioid supply and drug overdose deaths in US counties. *Addiction*, 118(6), 1072–1082. <https://doi.org/10.1111/add.16123>
16. Johnson, E. M., Lanier, W. A., Merrill, R. M., Crook, J., Porucznik, C. A., Rolfs, R. T., & Sauer, B. (2013). Unintentional prescription opioid-related overdose deaths: Description of decedents by next of kin or best contact, Utah, 2008–2009. *Journal of General Internal Medicine*, 28(4), 522–529. <https://doi.org/10.1007/s11606-012-2225-z>

17. Addiction to prescription painkillers and heroin surges in Pittsburgh, most widely abused drugs city-wide; – heroin causing more than half of overdose deaths; resources on private office-based treatment available for Pittsburgh consumers –. (2006, May 31). *PR Newswire*. <https://www.proquest.com/wire-feeds/addiction-prescription-painkillers-heroin-surges/docview/451189838/se-2>
18. Patel, K. B., & Rushefsky, M. E. (2021). *The opioid epidemic in the United States: Missed opportunities and policy failures* (1st ed.). Routledge. <https://doi.org/10.4324/9781003215899>
19. Freskos, B. (2013, January 22). Lawmakers, advocates work to mitigate prescription for disaster. *McClatchy - Tribune Business News*. <https://www.proquest.com/wire-feeds/lawmakers-advocates-work-mitigate-prescription/docview/1271956234/se-2>
20. Reyes, R. (2010, October 31). Pain clinics feeding habit. *McClatchy - Tribune Business News*. <https://www.proquest.com/wire-feeds/pain-clinics-feeding-habit/docview/761207402/se-2>
21. Schroeder, R. D., Ford, J. A., & Higgins, G. E. (2019). Death and destruction, hope and resilience. In *The prescription drug problem: A new American crisis?* (pp. 153–176). Praeger. <https://doi.org/10.5040/9798216000624.ch-005>
22. Zhao-Fleming, H., Hand, A., Zhang, K., et al. (2018). Effect of non-steroidal anti-inflammatory drugs on post-surgical complications against the backdrop of the opioid crisis. *Burns & Trauma*, 6. <https://doi.org/10.1186/s41038-018-0128-x>
23. Bart, G. (2012). Maintenance medication for opiate addiction: The foundation of recovery. *Journal of Addictive Diseases*, 31(3), 207–225. <https://doi.org/10.1080/10550887.2012.694598>
24. Mahase, E. (2019). Most fatal opioid overdoses are not caused by prescription drugs, US study finds. *BMJ*, 367, l5993. <https://doi.org/10.1136/bmj.l5993>
25. Martin, J., Cunliffe, J., Décary-Héту, D., & Aldridge, J. (2018). Effect of restricting the legal supply of prescription opioids on buying through online illicit marketplaces: Interrupted time series analysis. *BMJ*, 361, k2270. <https://doi.org/10.1136/bmj.k2270>