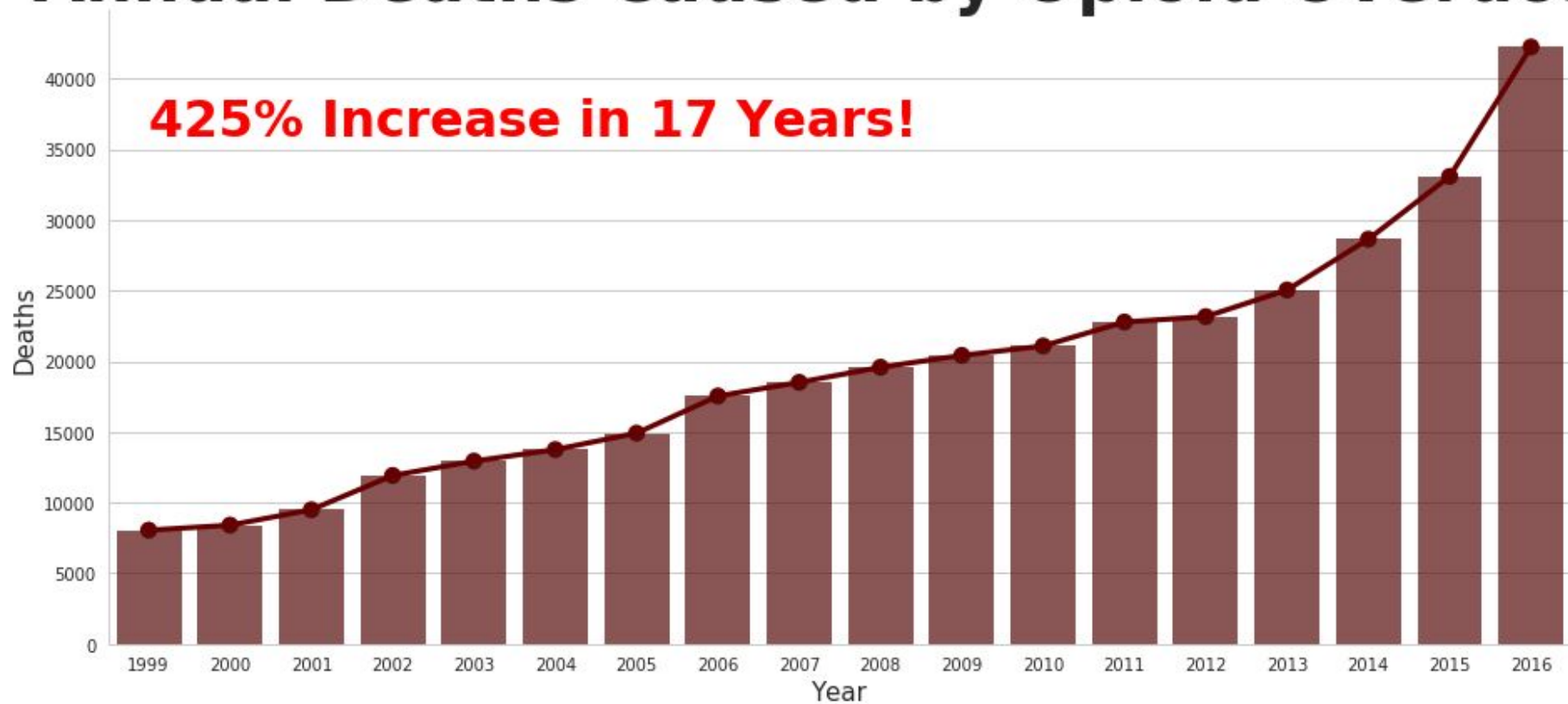


Predicting Opioid Prescription Rate By County

Daniel Lee

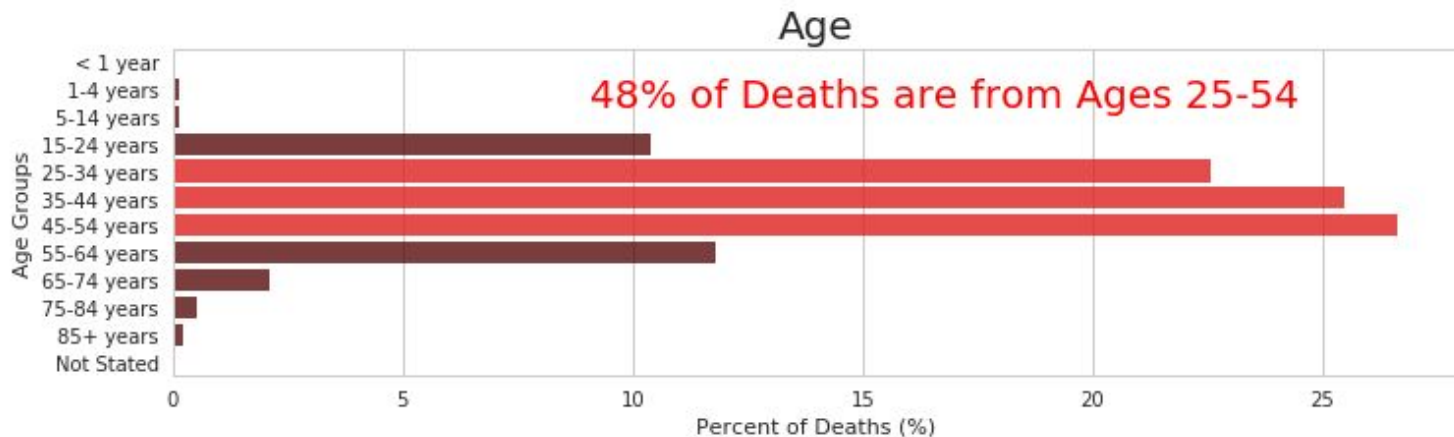
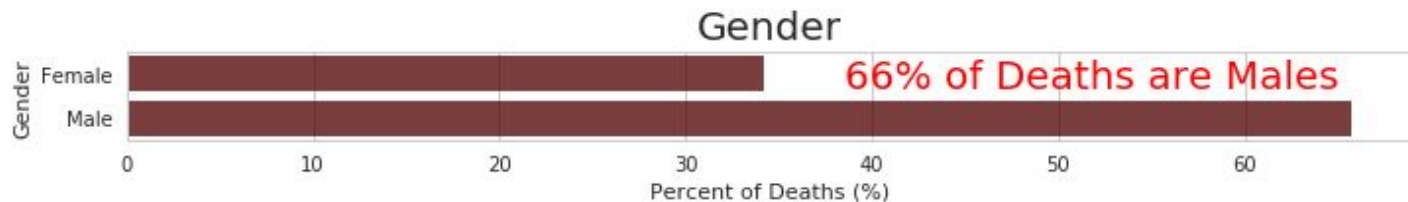
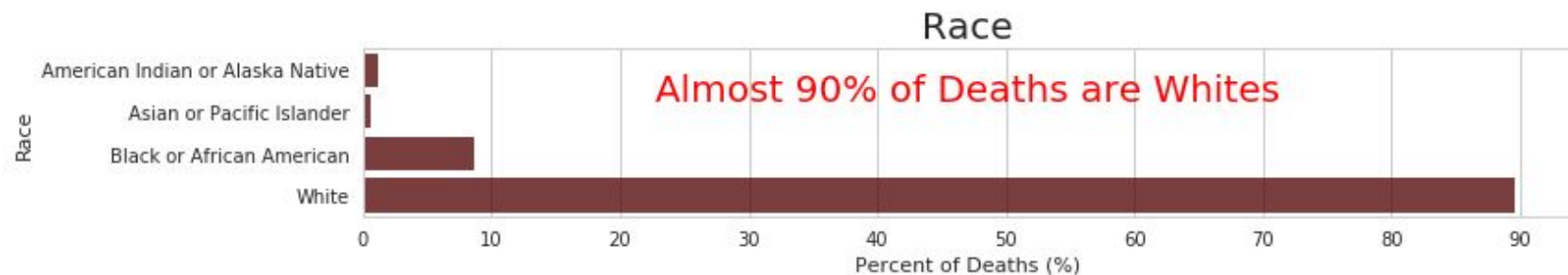
Annual Deaths Caused by Opioid Overdose



From 1999 to 2016 in America, the increase in deaths due to opioid overdose increased by 425%, from 8,050 deaths in 1999 to 42,249 deaths in 2016. That's over four times the increase in 17 years.

Who are the Victims?

Number of Deaths by Demographic Factors 1999 - 2016



Living With An Opioid Addiction



Why We Should Care

- Not represented in the death numbers are those who are struggling with opioid addiction
- Addiction to opioids, like any other addictions, leads to
 - Financial problems
 - Health problems
 - Lack of self control
 - Relational difficulties
 - Destruction of every aspect of the person
 - End result is death
- Dangers recognized by other nations
 - E.g. in the 1950s China
 - Banned opium
 - Executed those who were caught selling opium

What is the opioid epidemic?

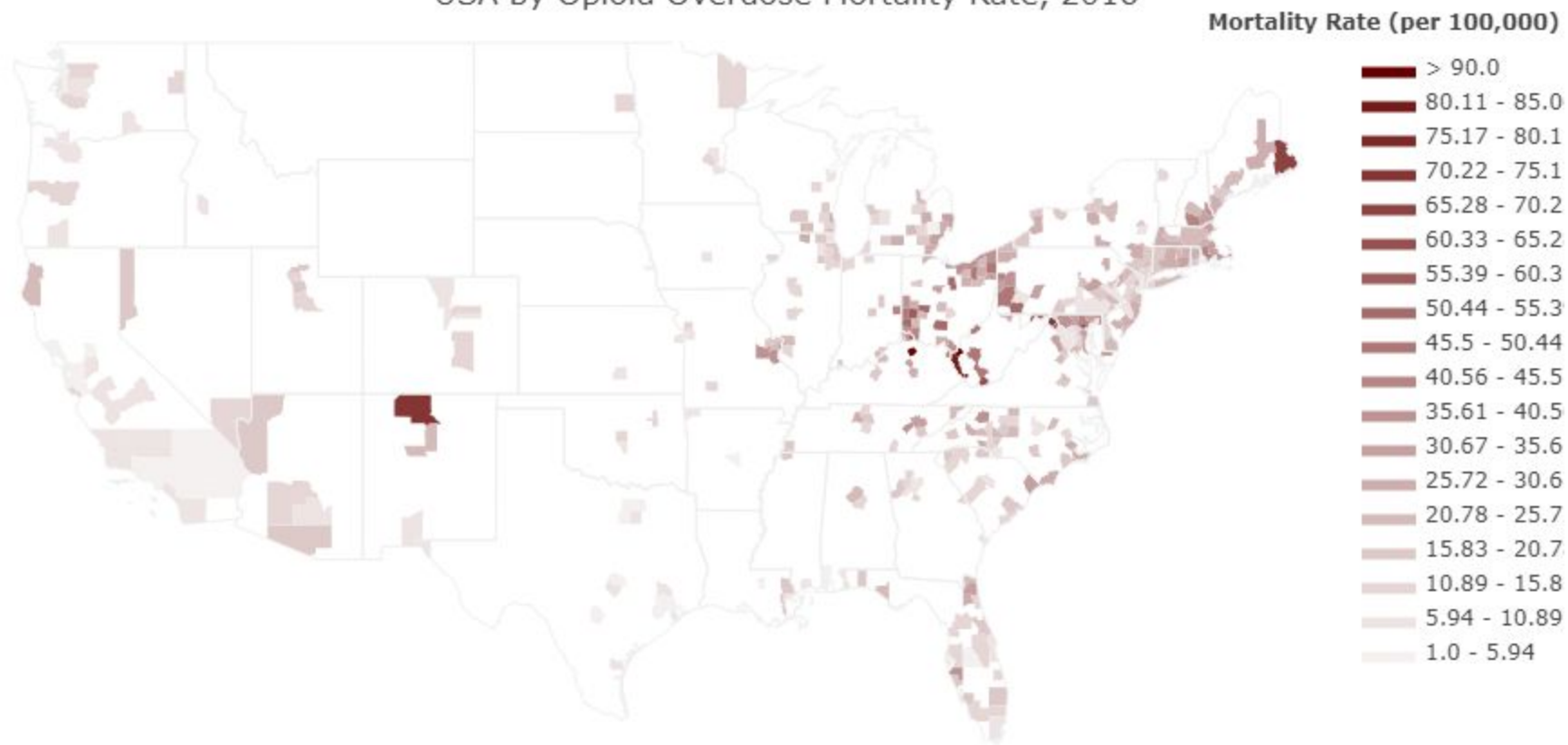
- Wikipedia defines the opioid epidemic as the following:
 - Rapid increase in the use of prescription and non-prescription opioid drugs in the United States and Canada beginning in the late 1990s and continuing throughout the next two decades.
- Opioids
 - Diverse class of moderately strong painkillers
 - Oxycodone (commonly sold under the trade names OxyContin and Percocet)
 - Hydrocodone (Vicodin)
 - Fentanyl
 - Synthesized to resemble other opiates such as opium-derived morphine and heroin

Source: https://en.wikipedia.org/wiki/Opioid_epidemic

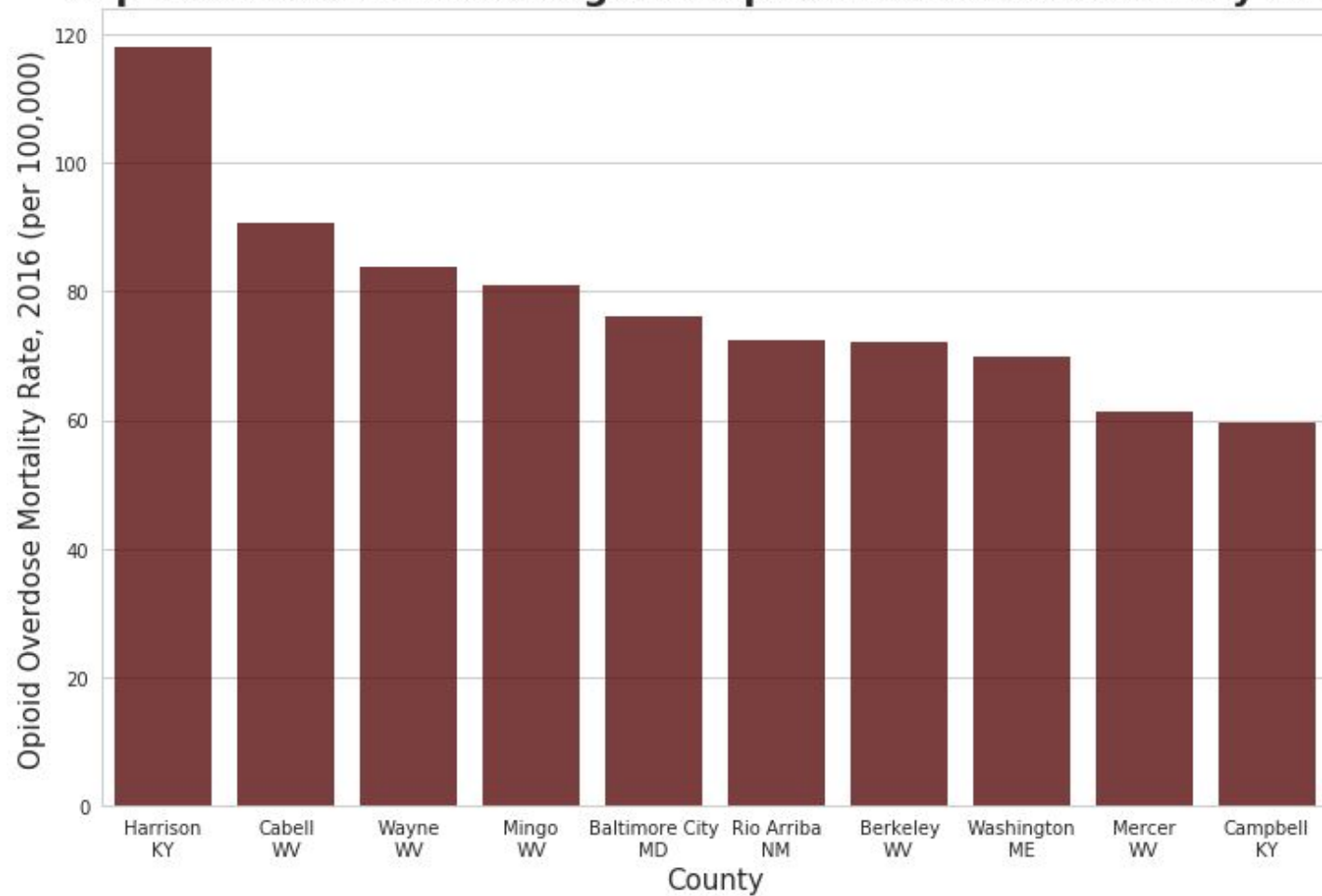
Problem

- Predict the county's mortality rate caused by drug overdose for 2016 given various independent variables, such as drug overdose for 2016 given a county's median household income, population estimates, unemployment, poverty rate estimates, educational attainment, and opioid prescription rate.

USA by Opioid Overdose Mortality Rate, 2016



Top Ten Counties with Highest Opioid Overdose Mortality Rate



Counties with High Mortality Rate Due to Opioid Addiction

- Highest county: **Harrison, KY**
 - 118 deaths per 100,000 people due to opioid overdose
 - 0.118% of all the people in county dying because of opioid overdose
- **West Virginia**
 - Five counties in the top ten highest mortality rate

Description of Data

- US County Opioid Prescribing Rates 2016
 - <https://www.cdc.gov/drugoverdose/maps/rxcounty2016.html>
- Unemployment, median household income, educational attainment, poverty rate estimates, and population estimates:
 - <https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/>
- Underlying Cause of Death
 - https://raw.githubusercontent.com/danielhanbitlee/Springboard/master/capstone_project/drug_overdose_death_opioid_2016.txt
 - <https://wonder.cdc.gov/controller/saved/D77/D39F042>
 - Note that statistics representing zero to nine deaths are suppressed at the region, state and county level.

Variables

- **_%_high_school_diploma_only_2012-2016** - Estimated percent of people with high school diploma only (%)
- **_%_povall_2016** - Estimated percent of people of all ages in poverty 2016 (%)
- **county** - County's name
- **county_state** - County's name and state the county is located in

Variables

- **crude_opioid_mortality_rate** - Estimated rate for deaths caused by opioid overdose in the county for 2016 (per 100,000 people)
 - Specifically, the types of drug-related deaths include the following:
 - Drug poisonings (overdose)
 - Unintentional
 - Suicide
 - Homicide
 - Undetermined

Variables

- **crude_opioid_mortality_rate**
 - The specific drugs included in the death rates are the following:
 - Opium
 - Heroin
 - Other opioids
 - Methadone
 - Other synthetic narcotics
 - Other and unspecified narcotics

Variables

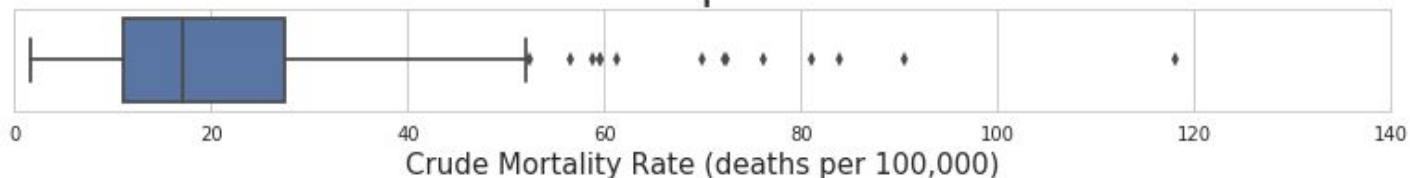
- **fips_code** - County's unique code
- **fips_padded_zeros** - County's unique code padded with zeros in front
- **gq_estimates_2016** - 7/1/2016 Group Quarters total population estimate (number of people)
 - all people not living in housing units (house, apartment, mobile home, rented rooms)
 - e.g. people living in correctional facilities
- **median_household_income_2016** - Estimate of median household Income, 2016 (in dollars)
- **opioid_prescription_rate** - Prescription rate for opioids for the county (per 100 people)
- **pop_chg_2016** - Net change in resident total population 7/1/2015 to 7/1/2016 (number of people)

Variables

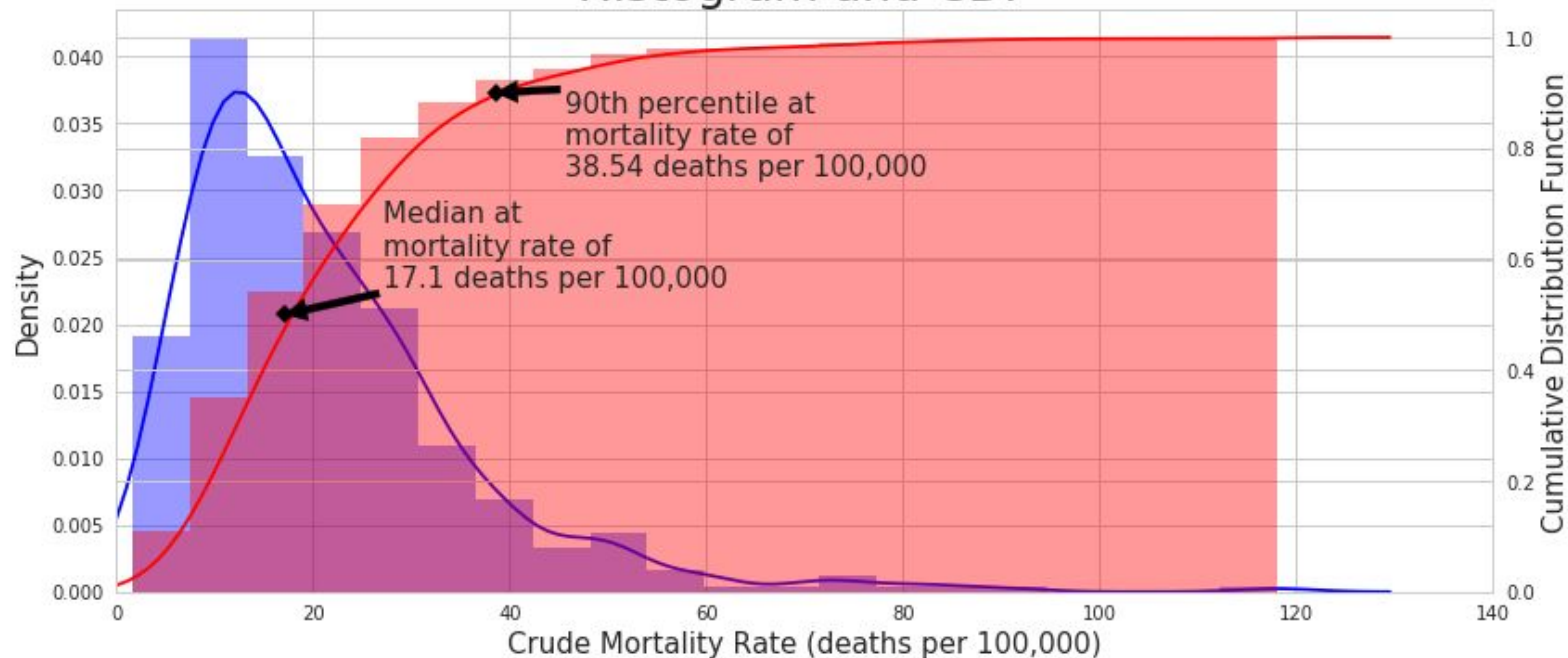
- **pop_estimate_2016** - Estimated population for the county in 2016
- **rate_birth_2016** - Birth rate in period 7/1/2015 to 6/30/2016 (per 1000 people)
- **rate_death_2016** - Overall death rate (all causes, not only caused by opioid overdose) in period 7/1/2015 to 6/30/2016 (per 1000 people)
- **rate_domestic_mig_2016** - Net domestic migration rate in period 7/1/2015 to 6/30/2016 (per 1000 people)
- **rate_international_mig_2016** - Net international migration rate in period 7/1/2015 to 6/30/2016 (per 1000 people)
- **state** - State the county is in
- **unemployment_rate_2016** - Unemployment rate, 2016 (%)

Distribution of County Level Opioid Overdose Mortality Rate, 2016

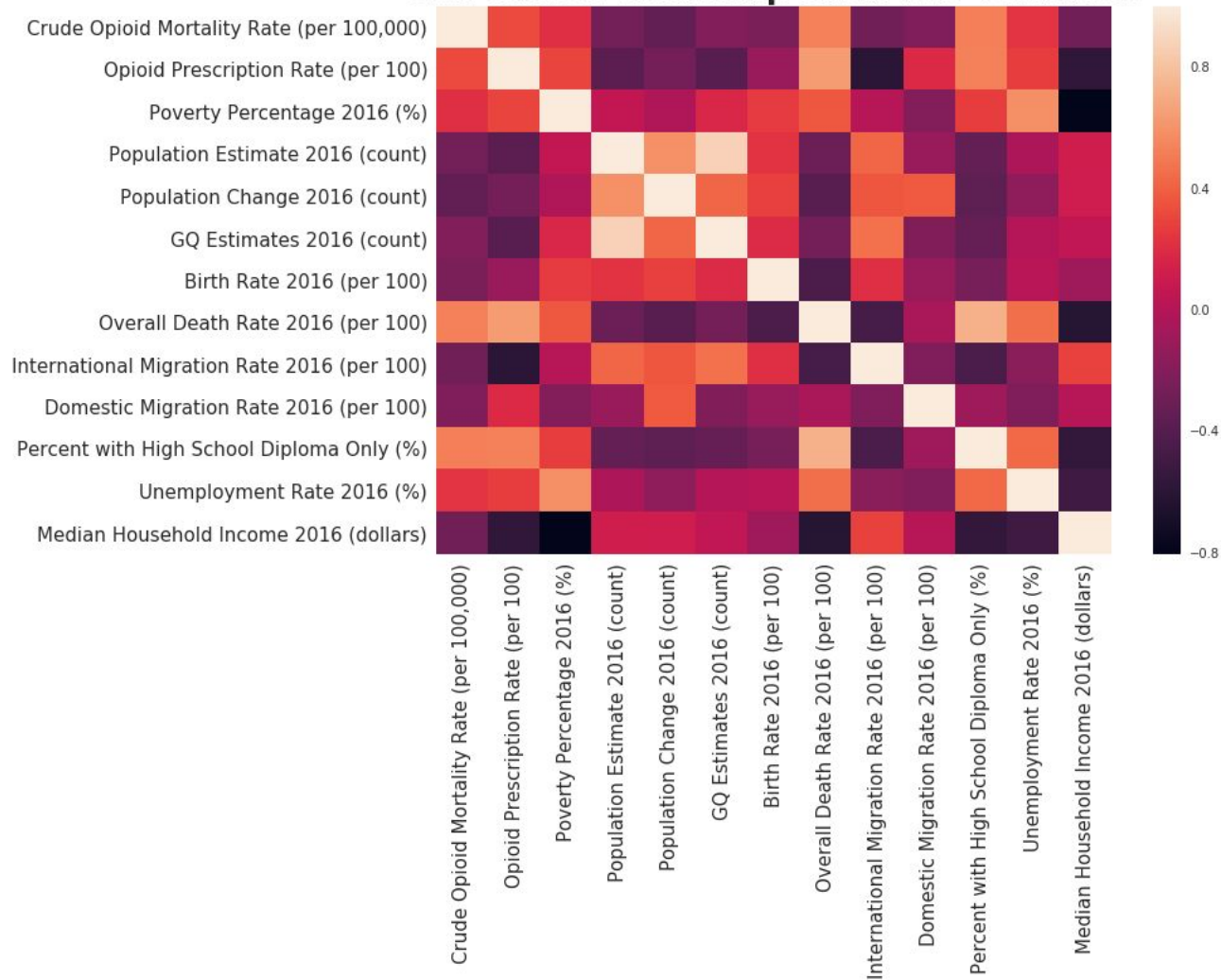
Boxplot



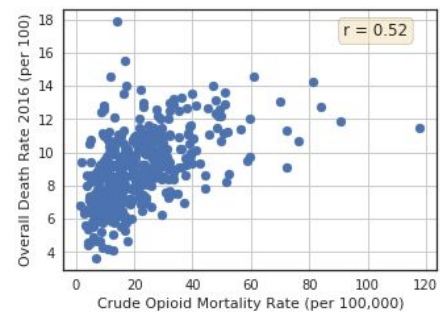
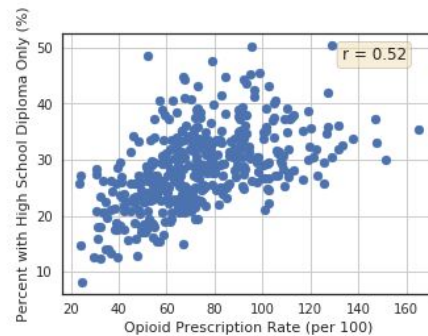
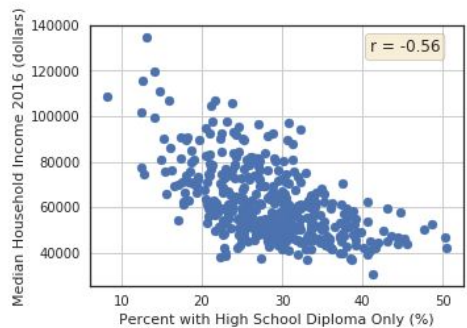
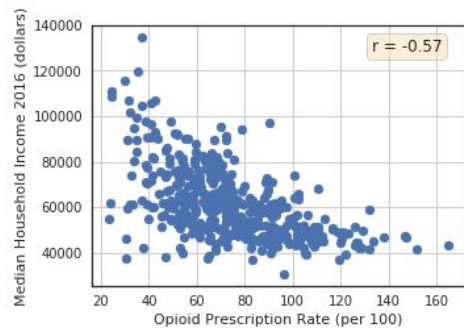
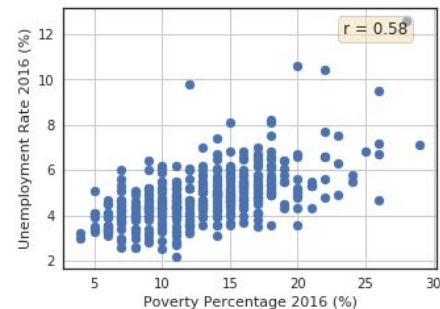
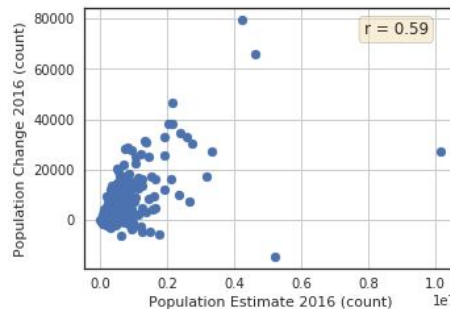
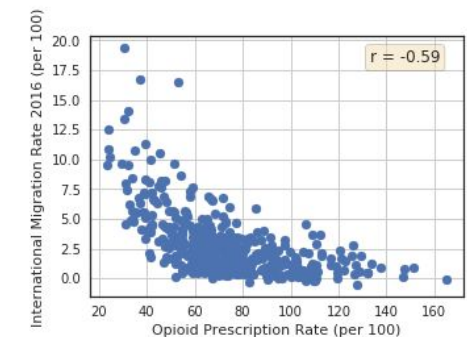
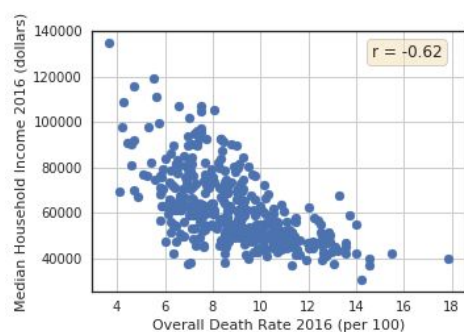
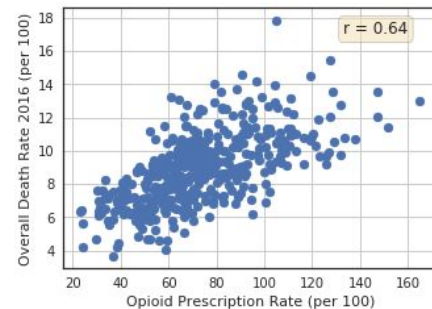
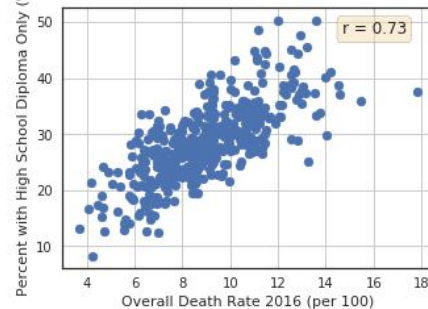
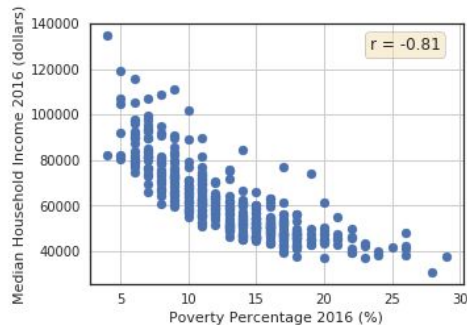
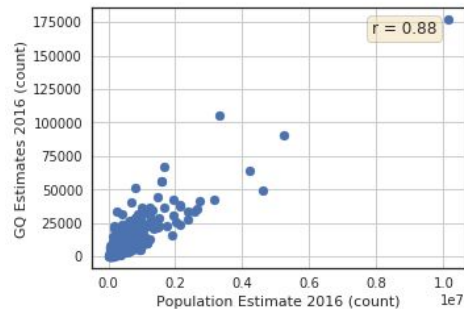
Histogram and CDF



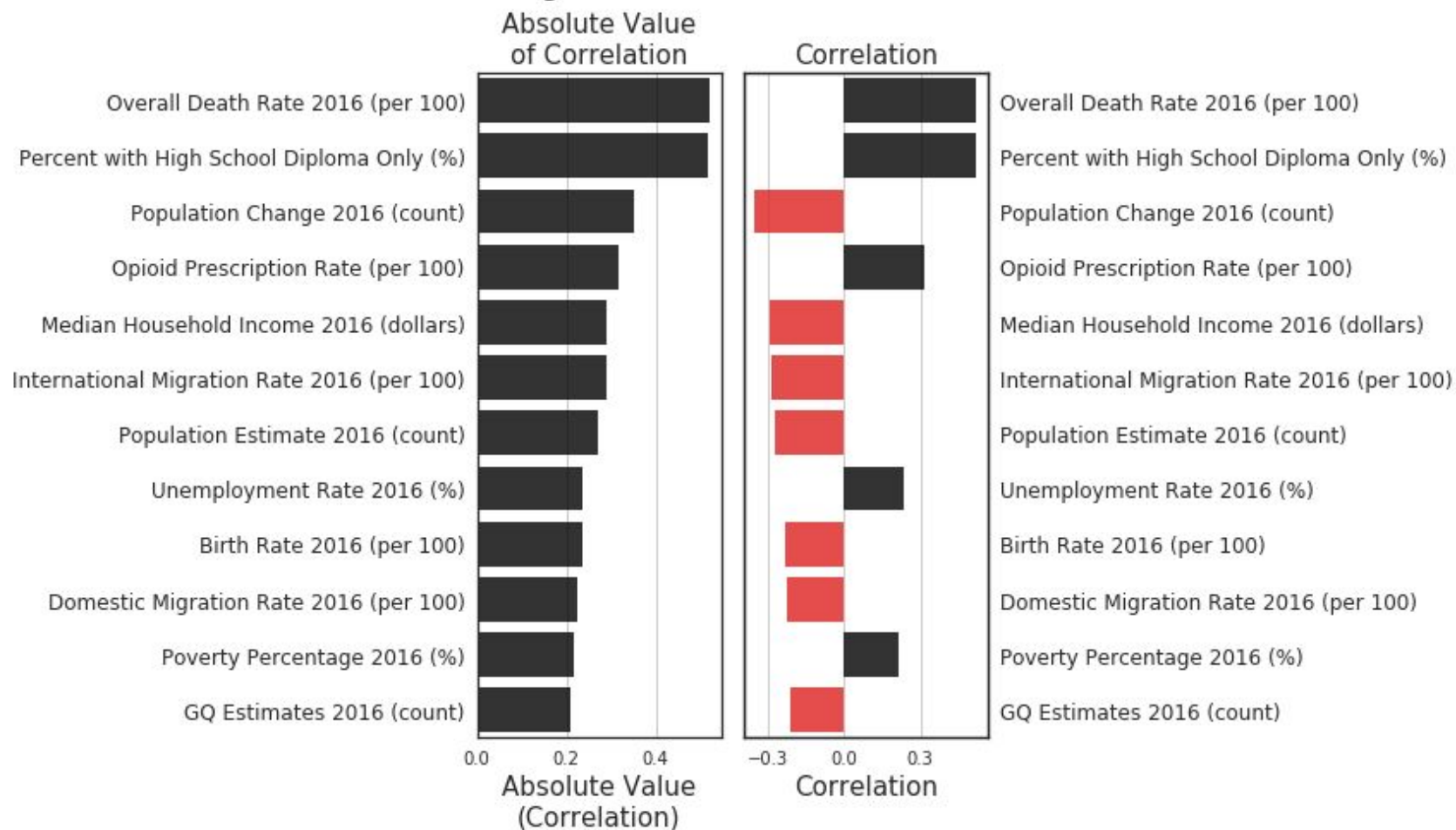
Correlation Heat Map of All the Variables



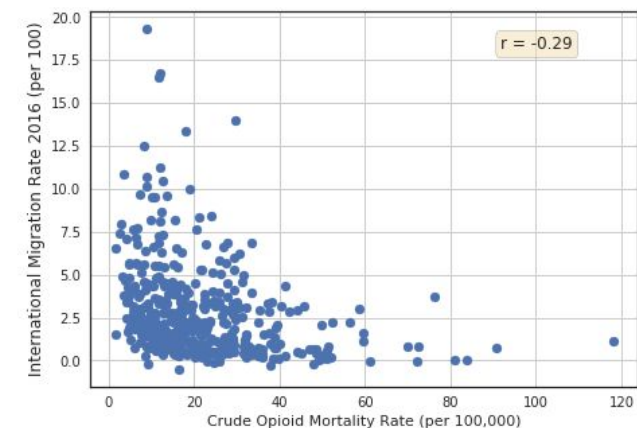
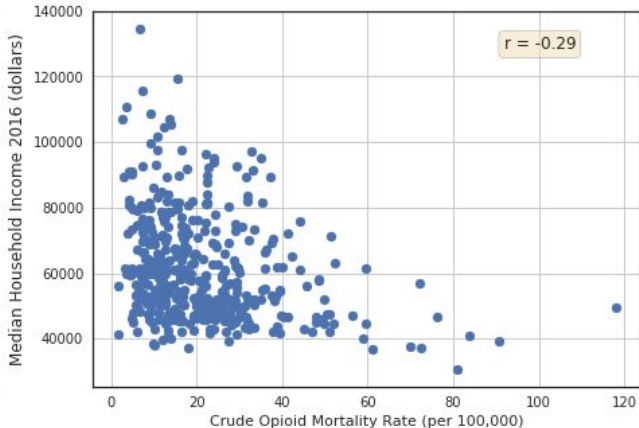
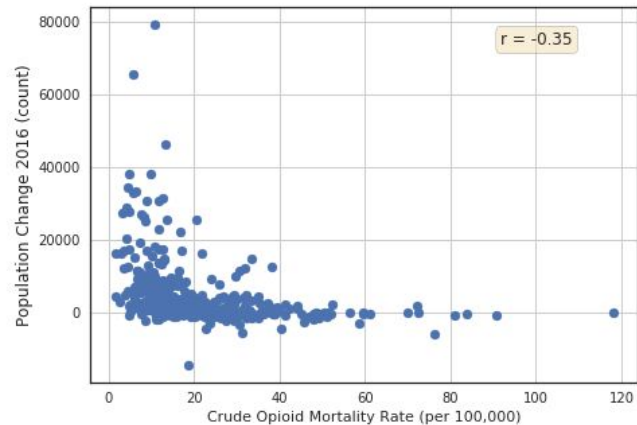
Top 12 Correlation Magnitude Variable Pairs



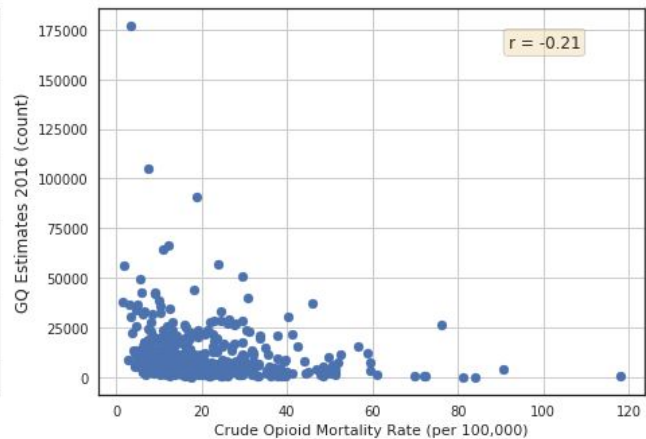
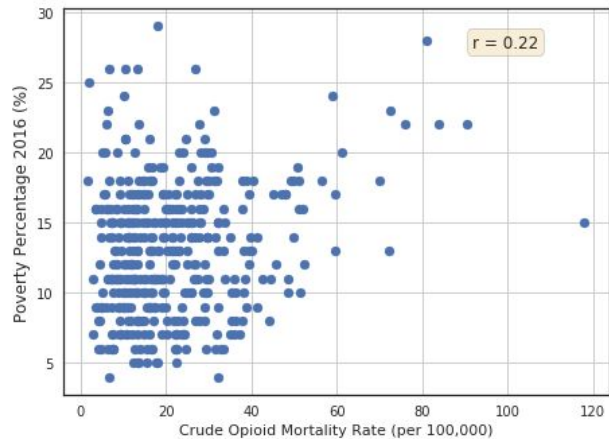
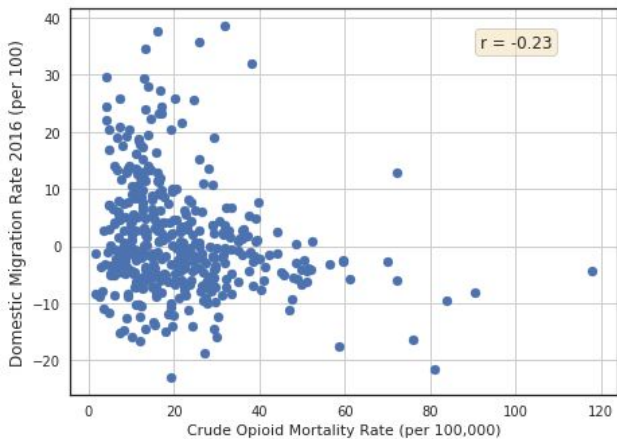
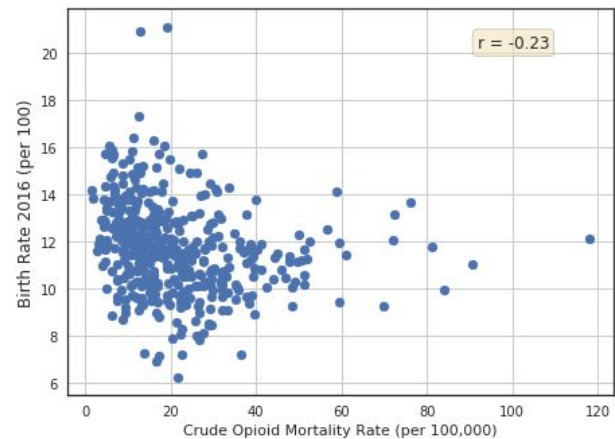
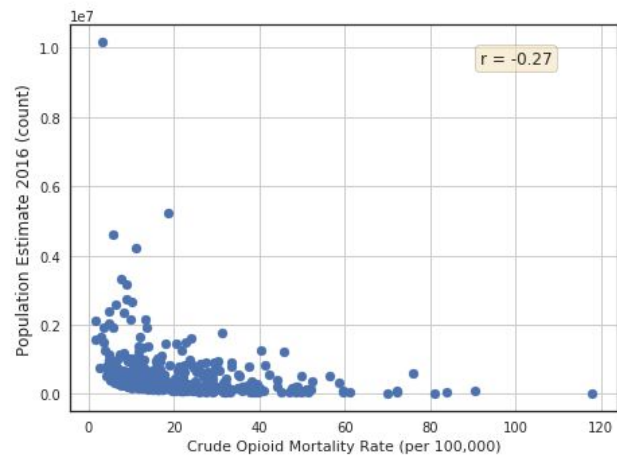
Pearson's Correlation Between Crude Mortality Rate and All Other Variables



Scatter Plots of Crude Opioid Mortality Rate and Other Variables



Scatter Plots of Crude Opioid Mortality Rate and Other Variables



Hypothesis Test of Correlation By Permuting Samples

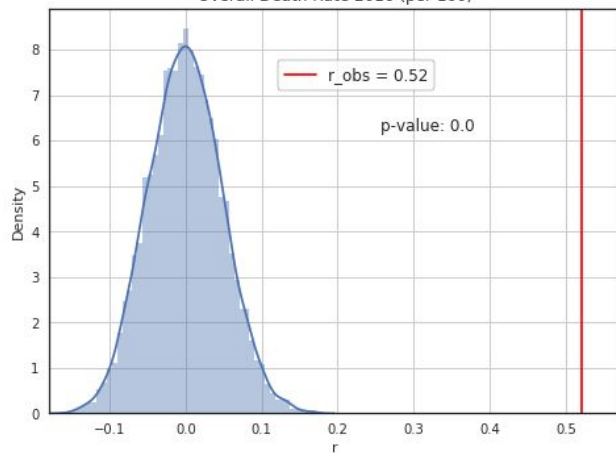
- Could these correlation values be due to chance?
- Or are the correlation values statistically significant?
- To test this, let's run a hypothesis test of correlation by permuting samples for Crude Opioid Mortality Rate (per 100,000) and all other continuous variables

Hypothesis Test of Correlation By Permuting Samples - Procedure

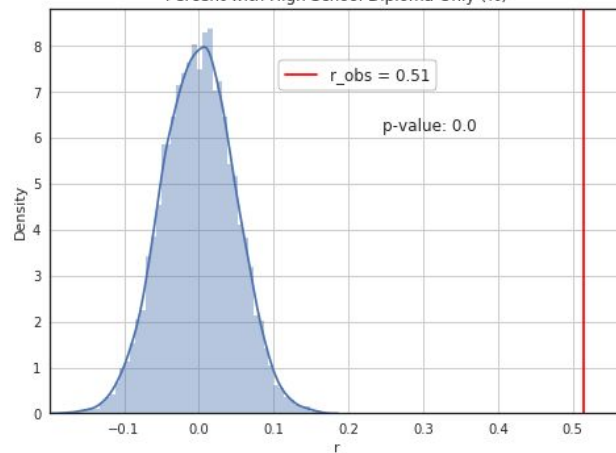
1. Null hypothesis: Crude Opioid Mortality Rate (per 100,000) and one of the twelve continuous variable are completely uncorrelated. That is, $\rho=0$.
2. Alternative hypothesis: Crude Opioid Mortality Rate (per 100,000) and the other variables are correlated. That is, $\rho\neq0$.
3. Simulate the data assuming null hypothesis is true by permuting the Crude Opioid Mortality Rate (per 100,000) column 10000 times and calculating the correlation with the other variables. This will create 10000 permuted replicates of sample Pearson correlation r .
4. Use sample Pearson correlation, rr , as a test statistic.
5. Compute p-value as fraction of permuted replicates that have r more extreme than the observed r .
6. Draw a conclusion based on significance level $\alpha=0.01$.

Hypothesis Test of Correlation By Permuting Samples: Distribution of Correlation Values

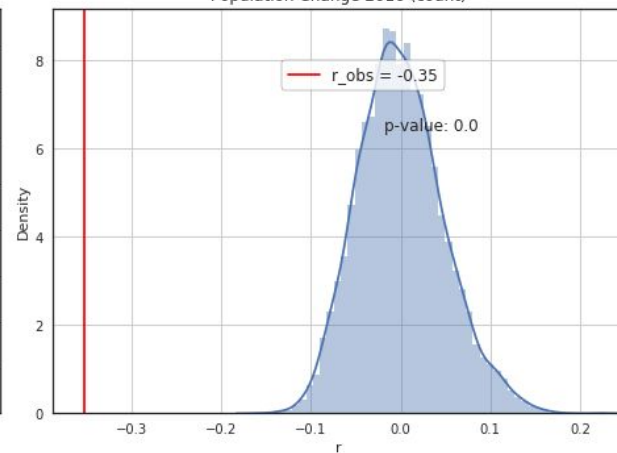
Opioid Mortality Rate (per 100,000) and
Overall Death Rate 2016 (per 100)



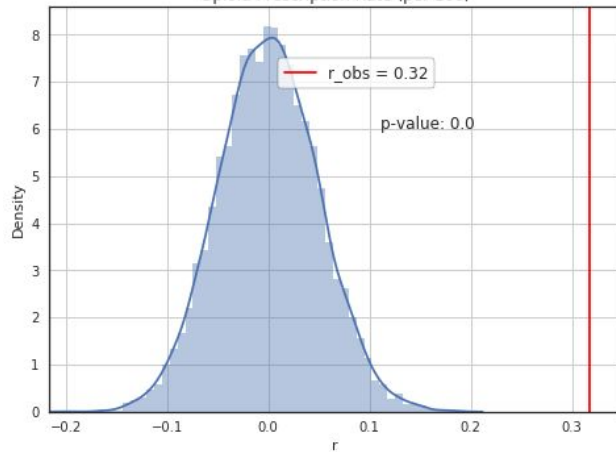
Opioid Mortality Rate (per 100,000) and
Percent with High School Diploma Only (%)



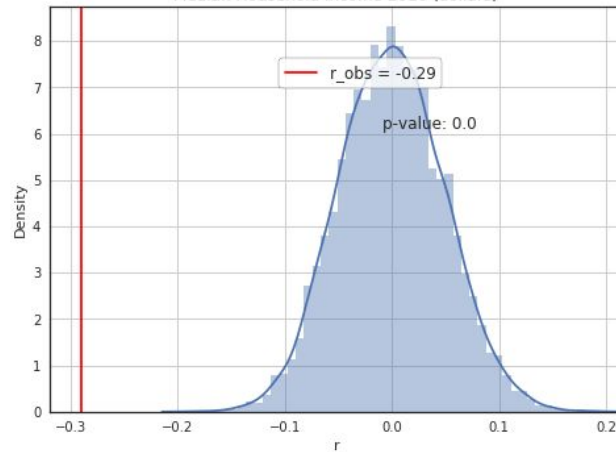
Opioid Mortality Rate (per 100,000) and
Population Change 2016 (count)



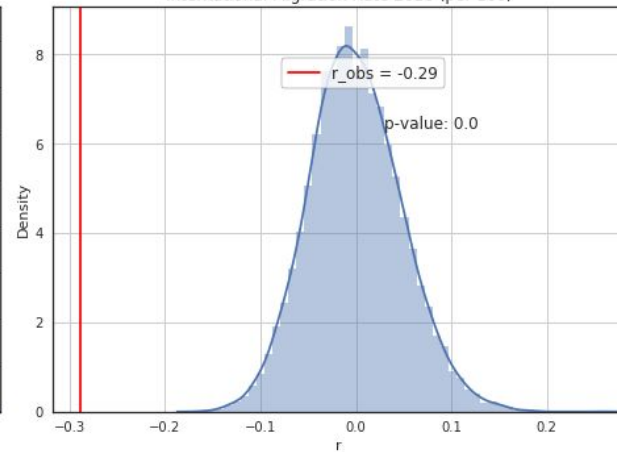
Opioid Mortality Rate (per 100,000) and
Opioid Prescription Rate (per 100)



Opioid Mortality Rate (per 100,000) and
Median Household Income 2016 (dollars)

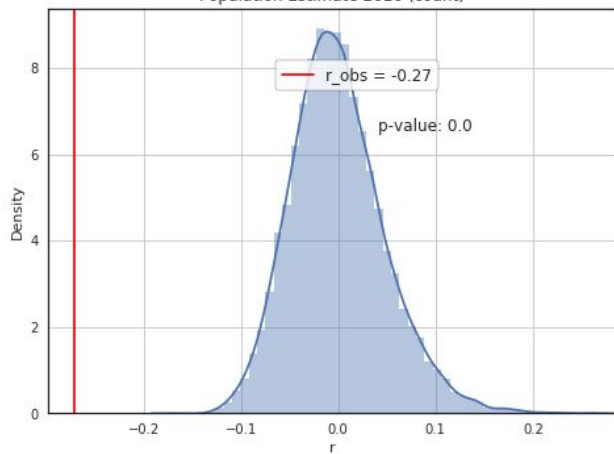


Opioid Mortality Rate (per 100,000) and
International Migration Rate 2016 (per 100)

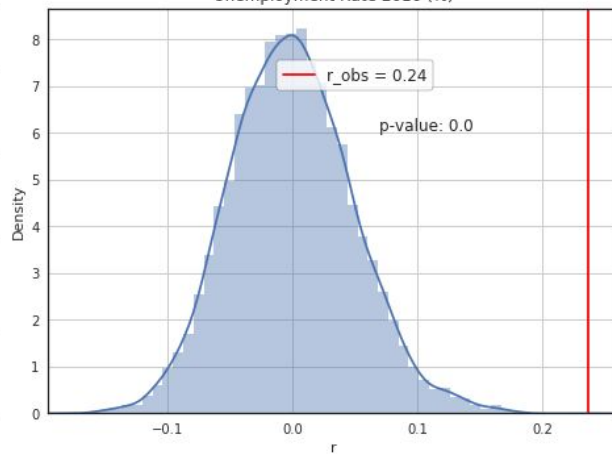


Hypothesis Test of Correlation By Permuting Samples: Distribution of Correlation Values

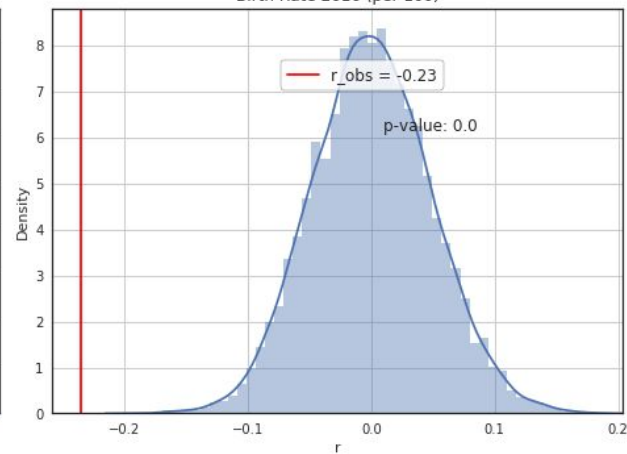
Opioid Mortality Rate (per 100,000) and
Population Estimate 2016 (count)



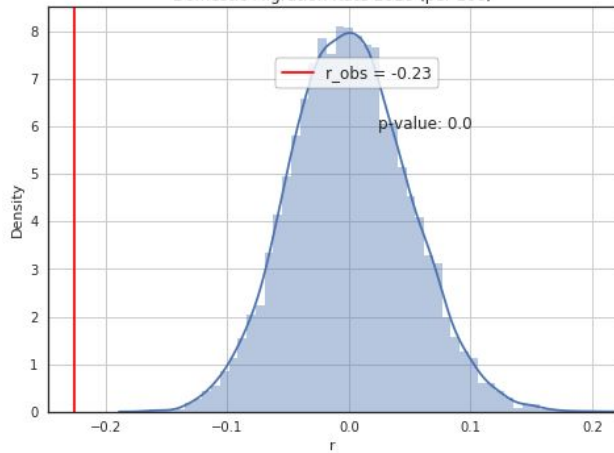
Opioid Mortality Rate (per 100,000) and
Unemployment Rate 2016 (%)



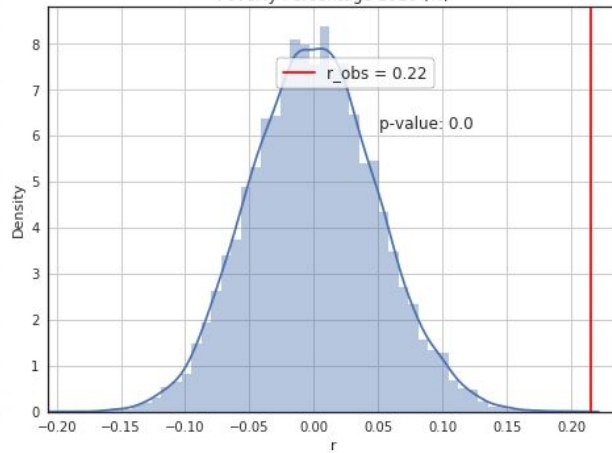
Opioid Mortality Rate (per 100,000) and
Birth Rate 2016 (per 100)



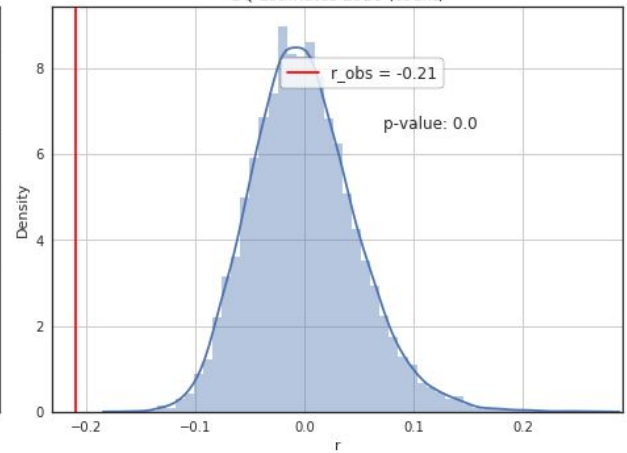
Opioid Mortality Rate (per 100,000) and
Domestic Migration Rate 2016 (per 100)



Opioid Mortality Rate (per 100,000) and
Poverty Percentage 2016 (%)



Opioid Mortality Rate (per 100,000) and
GQ Estimates 2016 (count)



Linear Regression and Confidence Interval Using Pairs Bootstrap

Let's examine the following pairs of variables to see how much one variable influences the other linearly:

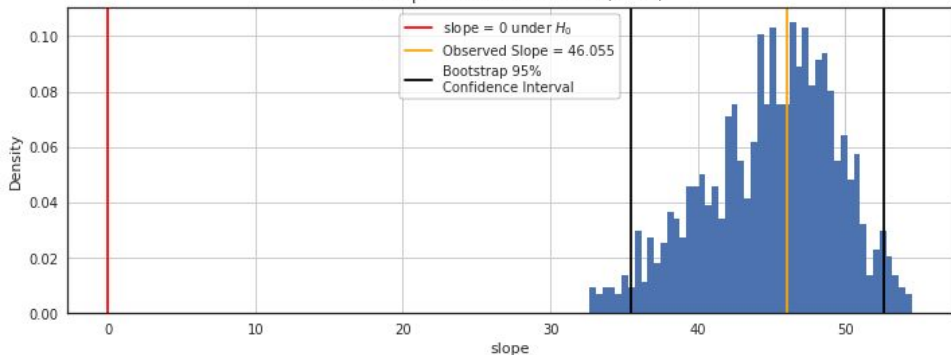
- GQ Estimates 2016 (count) vs Population Estimate 2016 (count) (independent variable vs independent variable)
- Percent with High School Diploma Only (%) vs Overall Death Rate 2016 (per 100) (independent variable vs independent variable)
- Overall Death Rate 2016 (per 100) vs Opioid Prescription Rate (per 100) (independent variable vs independent variable)
- Overall Death Rate 2016 (per 100) vs Crude Opioid Mortality Rate (per 100,000) (independent variable vs dependent variable)

Linear Regression and Confidence Interval Using Pairs Bootstrap - Procedure

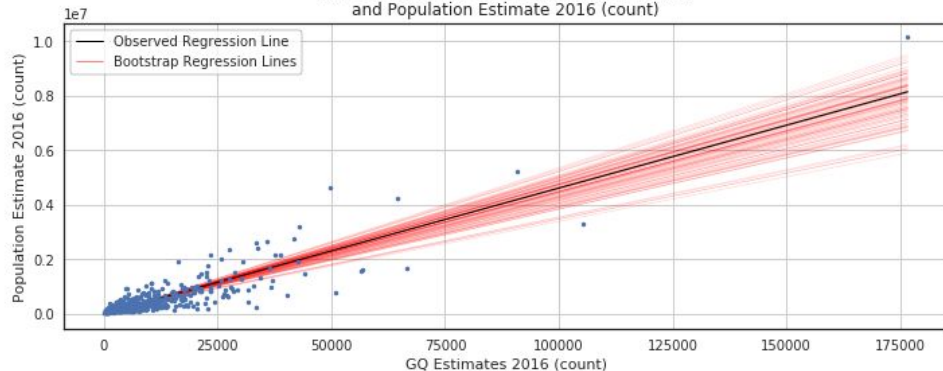
1. Null hypothesis is that the slope between the pairs of variables is zero in a linear regression model.
2. Sample (x, y) pair of sample size n with replacement from the original sample of size n to obtain bootstrap sample of pairs of size n .
3. Fit a linear regression model from the bootstrap sample of pairs in step 1 and calculate the slope and intercept.
4. Repeat steps 1 and 2 B times. Here, I will use $B=1000$. This will create a bootstrap sampling distribution of the slopes and intercepts.
5. Construct 95% bootstrap confidence interval using the bootstrap sampling distribution of the slopes.
6. If the confidence interval from step 4 does not contain number 0, then reject the null hypothesis at $\alpha=0.05$. Otherwise, do not reject the null hypothesis.

Linear Regression and Confidence Interval Using Pairs Bootstrap

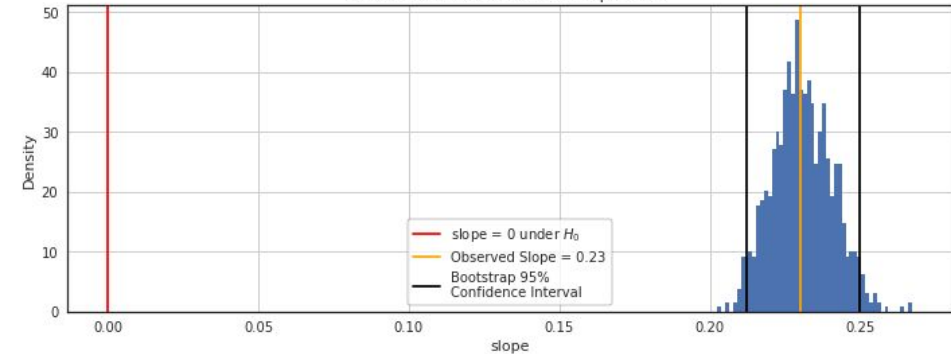
Bootstrap Sampling Distribution of the Slope of Regression Line for GQ Estimates 2016 (count) and Population Estimate 2016 (count)



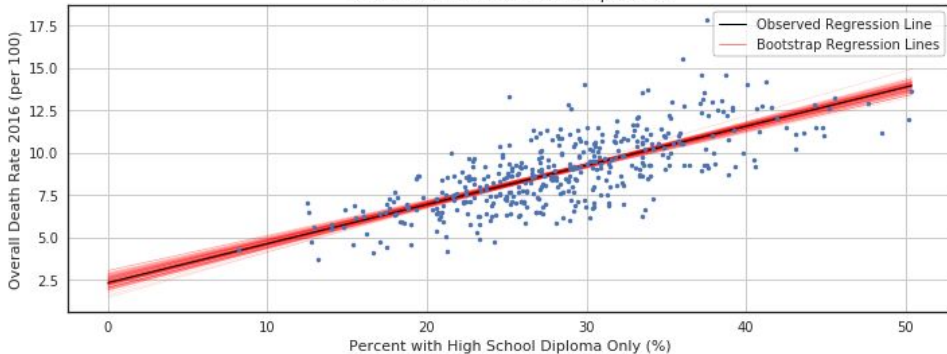
Observed Regression Line, Bootstrap Regression Lines and Scatter Plot of GQ Estimates 2016 (count) and Population Estimate 2016 (count)



Bootstrap Sampling Distribution of the Slope of Regression Line for Percent with High School Diploma Only (%) and Overall Death Rate 2016 (per 100)

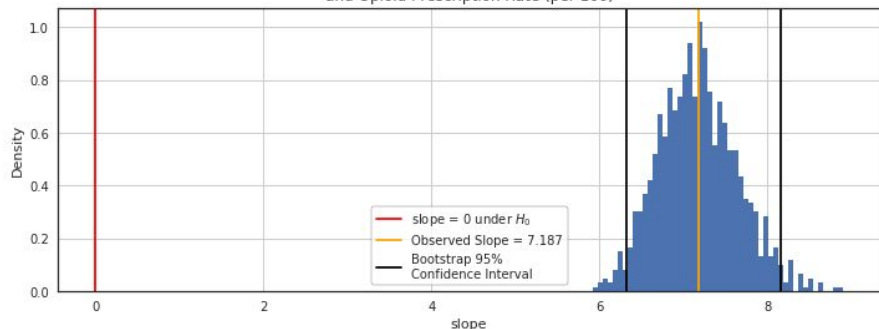


Observed Regression Line, Bootstrap Regression Lines and Scatter Plot of Percent with High School Diploma Only (%) and Overall Death Rate 2016 (per 100)

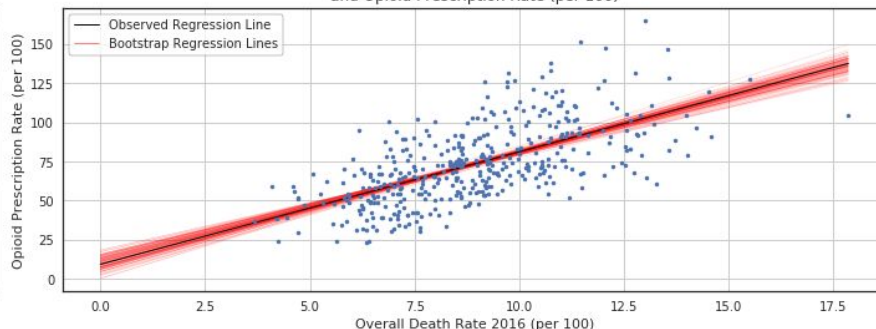


Linear Regression and Confidence Interval Using Pairs Bootstrap

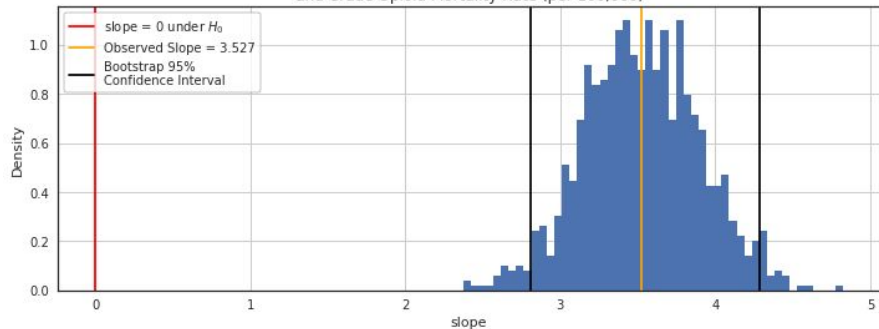
Bootstrap Sampling Distribution of the Slope
of Regression Line for Overall Death Rate 2016 (per 100)
and Opioid Prescription Rate (per 100)



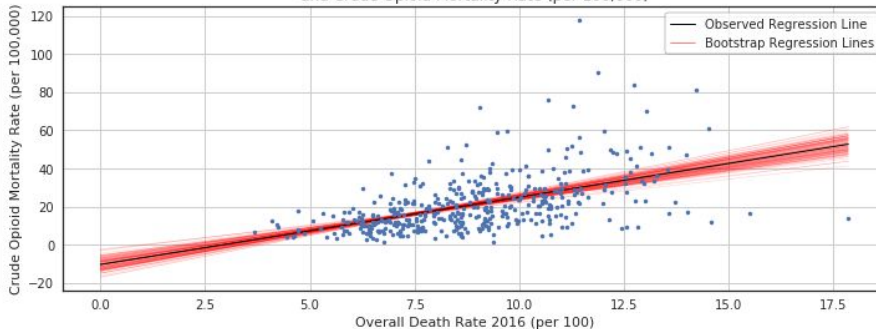
Observed Regression Line, Bootstrap Regression Lines
and Scatter Plot of Overall Death Rate 2016 (per 100)
and Opioid Prescription Rate (per 100)



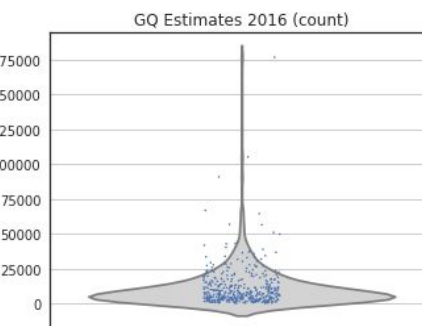
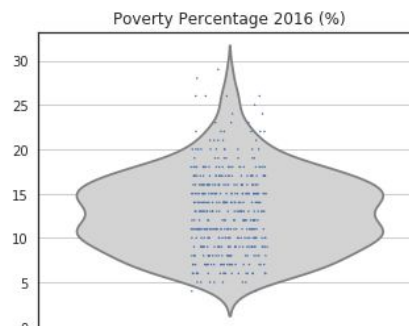
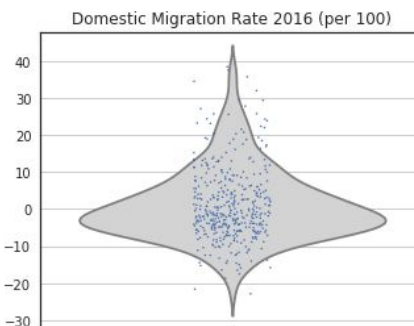
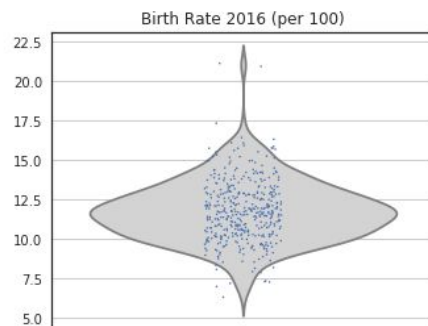
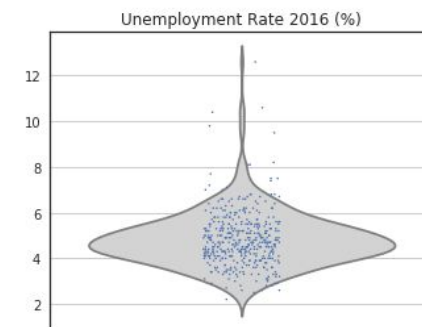
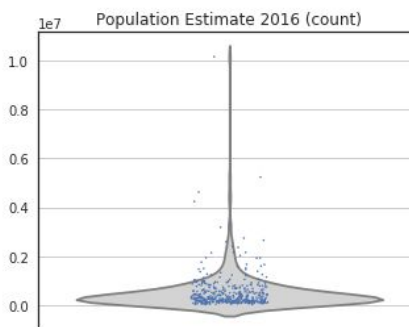
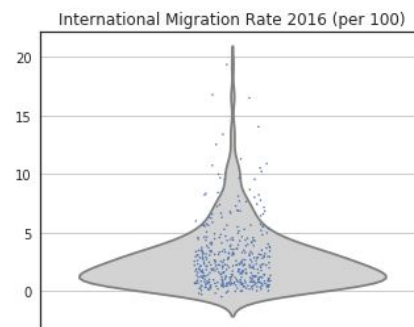
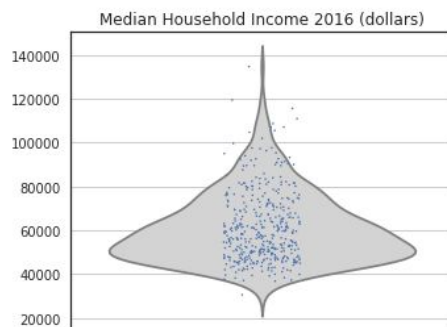
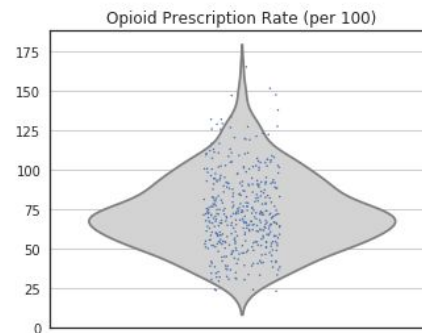
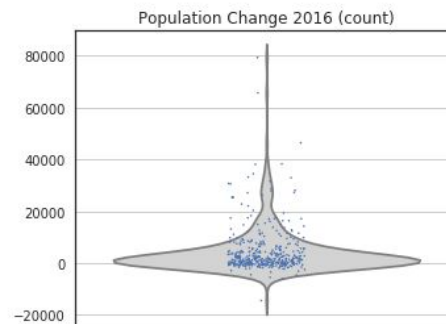
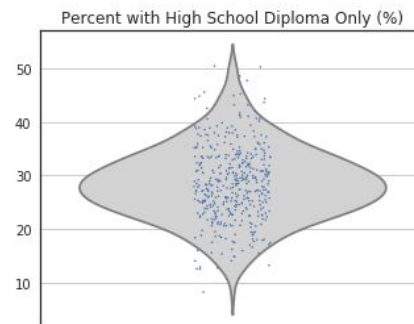
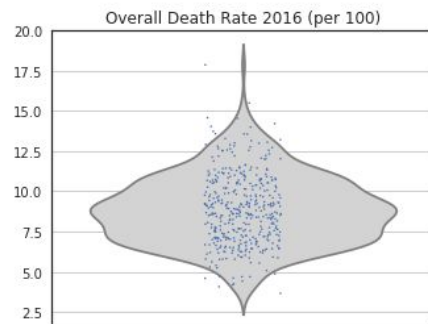
Bootstrap Sampling Distribution of the Slope
of Regression Line for Overall Death Rate 2016 (per 100)
and Crude Opioid Mortality Rate (per 100,000)



Observed Regression Line, Bootstrap Regression Lines
and Scatter Plot of Overall Death Rate 2016 (per 100)
and Crude Opioid Mortality Rate (per 100,000)



Violion Plots of All Variables Except Crude Mortality Rate



Conclusion From Exploratory Data Analysis

- Counties with higher opioid addiction mortality rates caused by opioid addiction have
 - Higher overall mortality rate
 - Have more adults with high school diplomas only
 - Higher opioid prescription rates
 - Decrease in population size in 2016
 - Lower median household income
- Possible reasons for these correlations
 - Population in these counties are older
 - These counties may not have good access to healthcare for the residents

Conclusion From Exploratory Data Analysis

- Further research
 - Why do counties that have high percentage of the population with only high school diplomas have high opioid overdose mortality rate?
 - Possibly population in these counties are older
 - Possibly people in these counties may not have access to higher education for some reason