Final Report

Erin Zheng & Tracy Cui

Topic: Does the legalization of marijuana have any effect on the crime rate of WA?

Data: https://www.ucrdatatool.gov/Search/Crime/State/RunCrimeStatebyState.cfm

1. Overview

- **What is your research question?**

Our research question is: Does the legalization of marijuana have any effect on the crime rate of WA?

We investigated the change in property crime rate and violate crime rate and total crime rate respectively.

- **Why is it important? What compelling policy problem will this research help us solve? To whom will your potential findings matter? What makes the problem "economics?"**

Since some states passed the legalization of marijuana, while some states didn't, we are wondering whether it's related to the crime rates or not. From this research, we hope to have an approximate idea of whether the legalization of marijuana is a good policy choice by the state government. The legalization of marijuana is an important public policy because it has potential effects on GDP and the employment rate of the economy, as well as the citizens' expectations on future economic development.

- **How does this question relate to economic theory or previous economic research?**

We found two previous economic research about the effects of legalization of Cannabis on the crime rate in Colorado and Washington State.

The first paper is by Lu, Ruibin, et al. Their results suggest that the legalization of recreational marijuana has had minimal effect on major crime in both Washington and Colorado State. They observed some immediate increases in crime at the point of legalization,

but no statistically significant long-term effects of recreational marijuana legalization or retail sales on violent or property crime rates.

The second paper is by Dragone and Davide, et al. They draw the conclusion that the concern that legalizing cannabis for recreational purposes may increase crime occupies a prominent position in the public debate about drugs is not justified. Instead of the possible increasing trend, the crime rate is actually dropping, which is a full decriminalization of drugs market. They also analyzed four main suspects about why legalizing the production and sale of cannabis affected criminal behavior.

Both of the research got the results that the legalization of marijuana has not resulted in a significant upward trend in crime rates.

- **Include at least TWO references of related economics literature.**

- The Cannabis Effect on Crime: Time-Series Analysis of Crime in Colorado and Washington State

  Lu, Ruibin, et al. "The Cannabis Effect on Crime: Time-Series Analysis of Crime in Colorado and Washington State." *Taylor & Francis,* https://www.tandfonline.com/doi/full/10.1080/07418825.2019.1666903.

- Crime and the Legalization of Recreational Marijuana

  Dragone, Davide, et al. "Crime and the Legalization of Recreational Marijuana." *Journal of Economic Behavior & Organization*, North-Holland, 20 Feb. 2018, https://www.sciencedirect.com/science/article/pii/S0167268118300386.

- **Describe the ideal dataset you would use to answer your question.**

The ideal dataset we want to find is the crime rate in WA by years from early 19th to now, with the information (by year) of population, total crime rate, violent crime rate, property crime rate (including detail category), and all the other factors that could've affected the crime rates. By including all the factors, we can completely remove OVBs. But it is too ideal because we don't know what could actually affect the crime rates, and some of the possible factors information like the education level for the state and the housing price level is hard to get. Thus, for our model, we will only consider the effect of the population and treatment effect on the crime rate.

2. Data

- **Describe the data you will use: what is the type of data. What is the unit of observation? What is the time period of the data?**

The data we used is a time-series data of estimated crime from 1960 to 2018 in Washington State specifically. The data has the information about the population and the crime rate of Washington state in each year.

The crime rate has been separated into Violent Crime and Property Crime. Violent Crime has been separated in Murder and nonnegligent manslaughter rate, Legacy rape rate, Robbery Rate and Aggravated assault rate. Property Crime has been separated in Burglary rate, Larceny-theft rate and Motor vehicle theft rate. All of the crime rate is calculated with per 100,000 population.

- **Who collected the data?**

The data we used with time period from 1960 to 2014 is collected by UCR, it is a Uniform Crime Reporting Program created originally by the FBI. It has four data collections: The National Incident-Based Reporting System (NIBRS), the Summary Reporting System (SRS), the Law Enforcement Officers Killed and Assaulted (LEOKA) Program, and the Hate Crime Statistics Program.

The data we add with time period from 2015 to 2018 is found from FBI UCS Annual Crime Reports, which is collected by the FBI.

- **Why this dataset and not others?**

This dataset is collected by the FBI, the representation of the crime institution of the US, which leads to a higher authenticity and accuracy of the data. This will decrease the bias and inaccuracy of our result draw based on this data set. Meanwhile, this data set includes all the variables we want to have for our ideal data.

- **Are there any limitations to the data?**

The data we use is actually find separately. We first found the data from 1960 to 2014 from UCR and plot a graph. From the graph, we found that we may want to have more recent

data after the legalization of marijuana in 2012 (not only 2013 and 2014). So, we went to the FBI web and found the following years' data from 2015 to 2018 separately on FBI annual Report. However, when we did that, we found that the data are not exactly the same even if they are from the same source. The differences are not really big but it may still cause some bias to our research.

- **What variables are you going to use from the data? Why?**

We are going to use the time series data with population, violent crime rate and property crime rate, since we want to focus more on broadly crime rate rather than specifically the category of the crime. We want to detect whether the legalization of marijuana has an effect on the total crime rate in WA.

Here is a summary of the variables that we used for our analysis:

Population	The population of Washington State
crimerate	The Crime Rate (per 100,000 population) of Washington State including both Property and Violent crime rates
Year	Year
Legalization	= 1 after 2012 when marijuana is legalized in Washington state; = 0 in and before 2012

- **What are the expected effects of the variables you have chosen?**

We expect the population and the crime rate is proportional to each other, and we expect that the legalization of marijuana has some effect on the crime rate of WA.

3. Summary Statistics

- **Provide a table of summary statistics**

	vars <dbl>	n <dbl>	mean <dbl>	sd <dbl>	median <dbl>	trimmed <dbl>
Year	1	59	1989.00	17.18	1989.0	1989.00
Population	2	59	4923328.86	1447243.46	4761000.0	4883393.57
ViolentCrimeRate	3	59	334.26	125.32	346.3	342.84
PropertyCrimeRate	4	59	4658.33	1249.70	4792.5	4711.48
crimerate	5	59	4992.59	1360.50	5108.4	5060.84

mad <dbl>	min <dbl>	max <dbl>	range <dbl>	skew <dbl>	kurtosis <dbl>	se <dbl>
22.24	1960.0	2018.0	58.0	0.00	-1.26	2.24
1945171.20	2853214.0	7535591.0	4682377.0	0.17	-1.36	188415.05
117.27	56.6	534.5	477.9	-0.67	-0.23	16.32
1466.29	2173.2	6646.6	4473.4	-0.37	-0.97	162.70
1624.63	2231.3	7113.0	4881.7	-0.41	-0.91	177.12

- **What is the mean and standard deviation?**

  For the crime rates, the mean is 4992.59, and the standard deviation is 177.12.

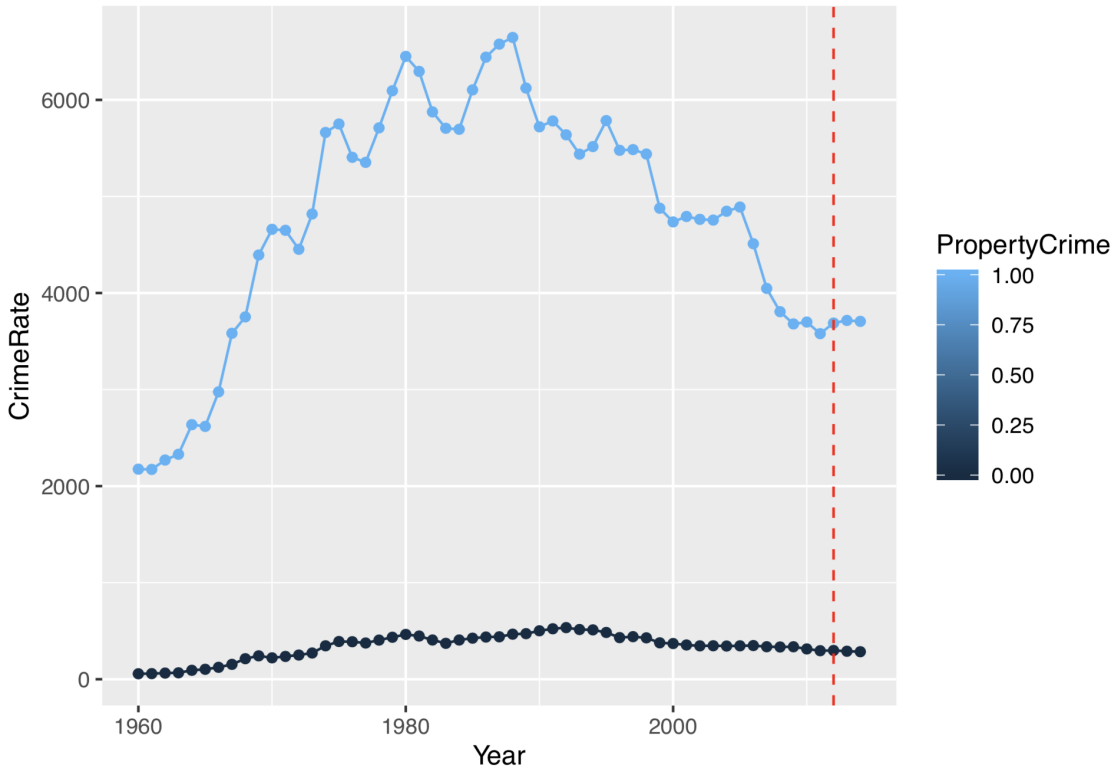
  For the Violent crime rates, the mean is 334.26, and the standard deviation is 16.32

  For the Property crime rates, the mean is 4658.33, and the standard deviation is 162.7

- **Provide a breakdown of the data? By country? gender? Racial composition? Etc.**

  Our data is specifically for Washington state, so we don't think we need to further break it down geographically, but instead, we broke down the data by the type of crime rates: Property Crime Rates and Violent Crime Rates.

- **Any charts or graphs that illustrate important features of the data.**



We plot the property crime rates and violent crime rates against time separately. From the graph, we can see that the property rate is more volatile and much higher than violent rate. After the legalization of marijuana in 2012, we can see a tiny decrease in violent crime rates and a tiny increase in property crime rates. But we couldn't tell whether the change is coming from or related to the legalization of marijuana. Also, in order to see the trend more clearly after 2012, we got more crime rate data after 2014 from FBI UCS Annual Crime Report.

With the new combined dataset, we got more plots:

- Violent Crime Rates:

- Property Crime rate:

- Total crime rates:

From these graphs, we can see that the violent crime rate has increased a little bit after the legalization in 2012, but the property crime has a decreasing trend after the legalization. Since the decreasing trend by property crime is larger than the little increase in violent crime, so the total crime rates have decreased after 2012.

4. Methodology

- **Set up an OLS model to answer your question.**

We thought about using difference-in-difference treatment effect OLS, but we noticed that property rate and property rates are not very parallel, and as mentioned before, property crime rates are more volatile than violent crime rates. Because of that, we decided to investigate the treatment effects for property and violent crime rates separately, so we set up regressions for each of them and set up another regression for total crime rates by summing property and violent crime rates.

For each regression, we used the same OLS model that includes the population, legalization, lag of crime rates, and year. Here is the population model:

$$\text{ViolentCrimeRate}_t = \beta_0 + \beta_1 Population_t + \beta_2 Legalization_t + \beta_3 Year + \beta_4 ViolentCrimeRate_{t-1} + u_t$$

$$\text{PropertyCrimeRate}_t = \beta_0 + \beta_1 Population_t + \beta_2 Legalization_t + \beta_3 Year + \beta_4 PropertyCrimeRate_{t-1} + u_t$$

$$\text{CrimeRate}_t = \beta_0 + \beta_1 Population_t + \beta_2 Legalization_t + \beta_3 Year + \beta_4 CrimeRate_{t-1} + u_t$$

- **Explain whether you expect the OLS assumptions to be valid.**

  The model is truly linear in parameters and independent variables

  1. The model is truly linear in parameters and independent variables

     → From our graph, we didn't see a clear linear relationship between crime rates and year. Instead, especially for the property crime rate, we saw a possible quadratic relationship. We are not sure whether including a lag of the crime rate will make it more linear, but at this point, we can only assume that because taking the square of years does not make too much sense.

  2. For any given independent variable X_{ji}, there must be variation in the values X_{j1}, X_{j2}, X_{j3}

     → for all the X's we are using, there are variations in the values → valid

  3. No perfect multicollinearity. There does not exist an exact relationship between the explanatory variables

     → Our explanatory variables are not related to each other → valid

  4. Zero conditional mean assumption

     $E(u|X_1) = 0$ and $E(u|X_1, X_2 \ldots X_k) = 0$

     → From our plots of the residuals, we can see that the errors are randomly distributed → valid

  5. Homoscedasticity

     $Var(u|X_1 \ldots X_k)\sigma_i^2 = \sigma_j^2$ for all $i \neq j$

     → We carried BP tests for all of our regressions and get p-values greater than 0.05, which means we fail to reject the null hypothesis for the BP test, so we expected no heteroscedasticity → valid

  6. No autocorrelation

$\mathrm{Cor}(u_i, u_j) = 0$

→ We carried BG tests for all of the regressions and get p-values greater than 0.05 which means we fail to reject the null hypothesis for the BG test, so we expected no heteroscedasticity → valid

7. The error term,u_i, is a random variable that is normally distributed with mean zero

   $E(u_i) = 0$

   → From our plots of the residual, we can see that the error terms are randomly distributed → valid

Overall, we expect all the OLS assumptions are mostly valid for our model, but there are some potential concerns and improvement that we should consider:

Our model might not be truly linear. We didn't carry White tests for testing for heteroscedasticity, which might detect more heteroscedasticity. The p-values for BG tests are low or right on the edge of failing to reject, so there are still concerns with possible autocorrelation for our model.

- **What forms of robustness or sensitivity checks could you perform to validate your results?**

We also got the Heteroskedastic-Robust Standard Errors for all of our estimations. we compared the standard errors of first stage regression with the Heteroskedastic-Robust Standard Errors, they are very different, which raised some concerns with the possible inefficiency of our estimations.

5. Results

- **Estimate and interpret your OLS results**
  - R output for Violent Crime Rate OLS model:

```
Call:
lm(formula = ViolentCrimeRate ~ Population + lag(ViolentCrimeRate,
    1) + Year + Legalization, data = Crime)

Residuals:
    Min      1Q  Median      3Q     Max
-42.685 -10.277   0.259  12.704  44.286

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)              -1.436e+04  4.721e+03  -3.042  0.00365 **
Population               -9.207e-05  2.774e-05  -3.319  0.00164 **
lag(ViolentCrimeRate, 1)  8.979e-01  3.621e-02  24.797  < 2e-16 ***
Year                      7.467e+00  2.446e+00   3.052  0.00355 **
Legalization              1.088e+01  1.162e+01   0.936  0.35338
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20.34 on 53 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.9736,    Adjusted R-squared:  0.9717
F-statistic: 489.6 on 4 and 53 DF,  p-value: < 2.2e-16
```

From the R output, we can find that legalization for violent is also not statistically significant, and it has a p-value of 0.35338, which is also way larger than 0.05, which means the legalization might not necessarily have effect on the violent crime rates. While the lag variable is significant at the 0.1% level, which means the crime rate is influenced by the change of the previous year. Also, variable population and year are both significant at the 1% level. Also, we got a really high R-squared value of 0.9717, which means our model can explain 97.17% of the change in violent crime rate.

  - R output for Property Crime Rate OLS model:

```
Call:
lm(formula = PropertyCrimeRate ~ Population + lag(PropertyCrimeRate,
    1) + Year + Legalization, data = Crime)

Residuals:
    Min      1Q  Median      3Q     Max
-520.82 -123.05   -7.43  143.73  568.39

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                -1.517e+05  6.631e+04  -2.287   0.0262 *
Population                 -1.013e-03  4.011e-04  -2.525   0.0146 *
lag(PropertyCrimeRate, 1)  8.710e-01  4.748e-02  18.343   <2e-16 ***
Year                       7.908e+01  3.442e+01   2.298   0.0256 *
Legalization               -3.182e+01  1.416e+02  -0.225   0.8230
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 249.2 on 53 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.961,     Adjusted R-squared:  0.958
F-statistic: 326.3 on 4 and 53 DF,  p-value: < 2.2e-16
```

From the R output, we can find that legalization is not statistically significant, and it has a p-value of 0.823, which is way larger than 0.05, which means the legalization might not necessarily have effect on the violent crime rates. While the lag variable is significant at the 0.1% level, which means the crime rate is influenced by the change of the previous year. Also, variable population and year are both significant at the 5% level. Also, we got a really high R-squared value of 0.958, which means our model can explain 95.8% of the change of property crime rate.

- R output for Crime Rate OLS model:

```
Call:
lm(formula = crimerate ~ Population + lag(crimerate, 1) + Year +
    Legalization, data = Crime)

Residuals:
    Min      1Q  Median      3Q     Max
-543.12 -131.90   -1.58  150.89  613.79

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)       -1.639e+05  6.852e+04  -2.392   0.0203 *
Population        -1.091e-03  4.132e-04  -2.641   0.0108 *
lag(crimerate, 1)  8.748e-01  4.508e-02  19.407   <2e-16 ***
Year               8.544e+01  3.556e+01   2.403   0.0198 *
Legalization      -1.947e+01  1.475e+02  -0.132   0.8955
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 259.1 on 53 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.9643,    Adjusted R-squared:  0.9616
F-statistic: 357.6 on 4 and 53 DF,  p-value: < 2.2e-16
```

With the total crime rate OLS model, we can find that legalization for crime rate is still not statistically significant, and it has a p-value of 0.8955, which is still way larger than 0.05, which means the legalization might not necessarily have effect on the violent crime rates. While all the other variables are all significant at the 5% level, which means the crime rate is influenced by the change of the previous year, population and years. For the total crime rate model, we still got really high R-squared value of 0.9616, which means our model can explain 97.96.16% of the change in crime rate.

- **Are the sign and magnitude of the estimates reasonable? Provide a perspective.**

From our data, we can see that the magnitude of the coefficients in our estimate are really small, which let us have some concern of the model. One of the possible reasons is our crime rate is calculated by per 100,000 population, which may cause our coefficient to become small.

The coefficient for population variable for all three model is negative. It is confusing since we think the crime rate will increase as the population increases. However, it does make sense because our crime rate is calculated with per 100,000 population. So, when the increasing percentage of crime happens is slower than the population increases will lead to a negative coefficient.

Even though the estimation might not be efficient, the coefficient estimation for legalization actually makes sense. The legalization will have a small positive effect on the violent crime rate because there might be more people got high and do something really crazy and bad without consciousness. The legalization has negative effects on property rate, which we had seen researches having the same results.

- **Are the results consistent with what you expect?**

From the estimation we found that the coefficient of legalization on all three models are not statistically significant, and the p-value are all larger than 0.05, which means all of the models reject our null that we think legalization has an effect on the crime rate. So the result is not really consistent with what we expect that legalization may have an effect on the crime rate in WA.

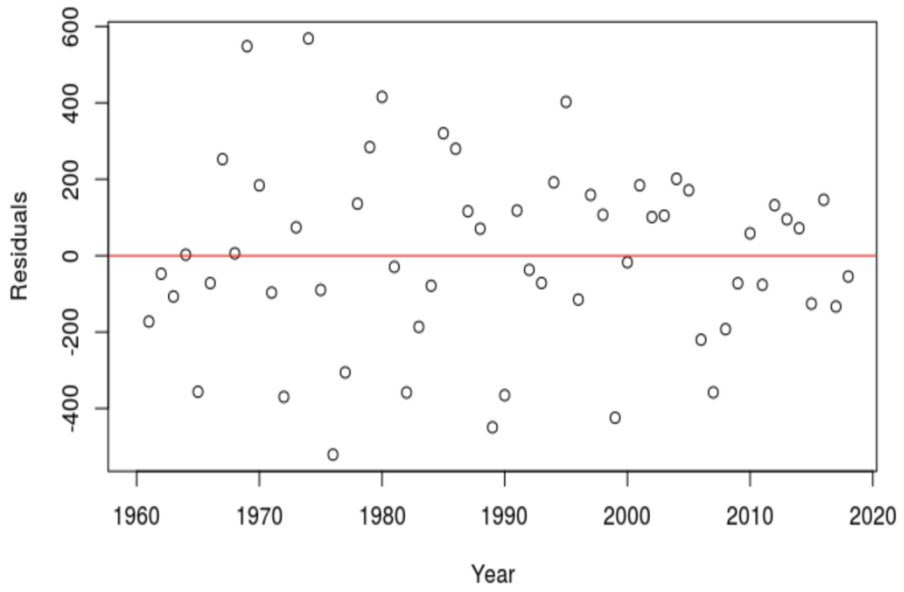
- **Discuss potential sources of bias (for your estimates and standard errors)**
  - Data may not perfectly accuracy for our research
  - Still a lot of other variables may influence the crime rate in WA

- **Conduct at least one test, that we have covered in class (i.e. an F-test for joint significance, a test for heteroskedasticity or autocorrelation if appropriate.)**

We conduct autocorrelation test for all three models since they are time-series data. We plot the graph of error term to see whether there is autocorrelation from the graph, and we did the BG test to confirm our thought.

- Violent Crime Rate:

  Graph:



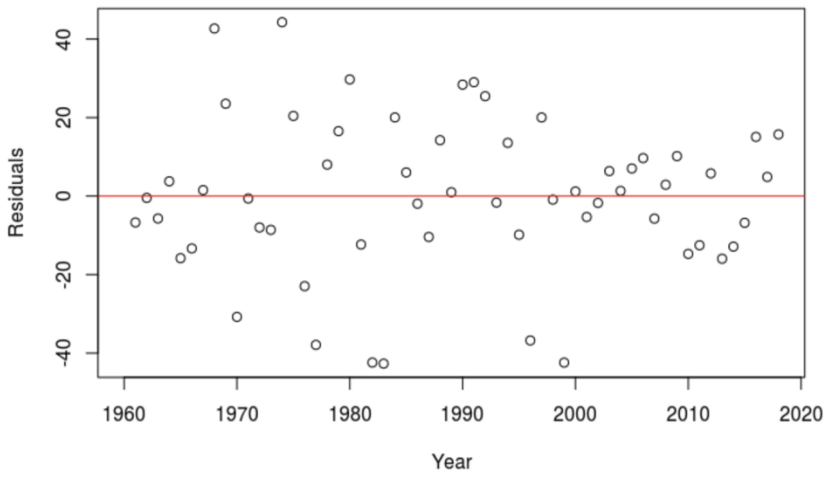
  BG test:

```
        Breusch-Godfrey test for serial correlation of order up to 1

data:  model1
LM test = 2.2126, df1 = 1, df2 = 52, p-value = 0.1429
```

- Property Crime Rate:

  Graph:



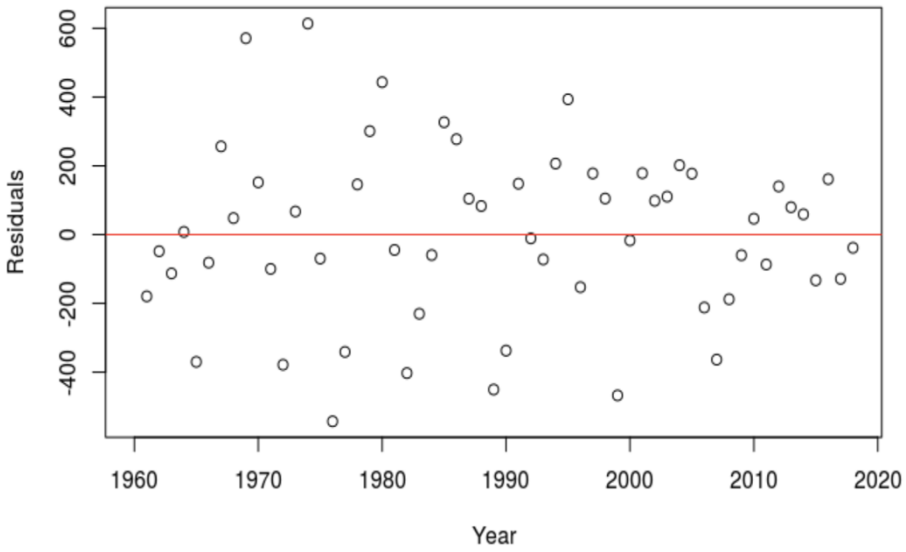
BG Test:

```
        Breusch-Godfrey test for serial correlation of order up to 1

data:  model2
LM test = 4.5131, df1 = 1, df2 = 52, p-value = 0.03841
```

- Total Crime Rate:

  Graph:



BG Test:

```
        Breusch-Godfrey test for serial correlation of order up to 1

data:  model3
LM test = 4.1099, df1 = 1, df2 = 52, p-value = 0.04777
```

6. Policy Implications, Discussion, or Conclusion

- **What future work should be done on this question? What extensions can be examined by future researchers?**

For the future, it would be valuable to do research on other states, not only WA. It would be interesting to make some comparison between WA and other states (recently legalized states), combined with more variables such as level of marijuana usage level, demand and

supply of marijuana. Also, we could probably do a difference-in-difference treatment effect between WA and CO so that we can have a more accurate estimation for WA.

From the graph, we find that it looks like a quadratic relationship between the crime rate and year, so maybe do more research on the quadratic relationship.

As for now, there are still concerns with omitted variable biases, because there might be more variables that could affect the crime rates, but we don't have access to those data right now. In the future, we could possibly include more variables into our model like the education levels, wage level (minimum wage), and house prices.

Also find better quality data. It would be better to have more data after legalization in 2012, since now only 7 years data after 2012. More data will help us better on the research of long-term effect by the legalization of marijuana on crime rates in different states.