Capstone Project - Hate Crimes in the United States

Benjamin Yang

2019/8/13

Introduction

A hate crime (also known as a bias-motivated crime or bias crime) is a prejudice-motivated crime which occurs when a perpetrator targets a victim because of their membership (or perceived membership) in a certain social group or race. In recent years, under various kinds of factors, the hate crimes in the U.S is surfacing a new trend. This short study will analyze all kinds of factors that may influence hate crimes and provide some reasonable explanation on it, by using OLS regression model and some interesting plots.

Variables and Data sources

Variables

```
Independent variables:
1. Median household income
2. Unemployment rate
3. Rate of people with only high school diploma
4. Rate of white people who are below poverty line
5. GINI index of each state
6. Rate of non-white people
7. Rate of non-citizen people

Additional independent variables:
8. Whether or not it is a union state or slave state during the civil war
9. Rate of LGBT people

Dependent variable:
10. Annual hate crimes per 100k population
```

In this research, we used data from United States Census Bureau for most of our variables (median household income, unemployment rate, high school accomplishment rate, white people poverty rate, GINI index, non-white rate) in 2016.
From Kaiser Family Foundation, we get the data for non-citizen rate in 2017.
From Federal Bureau of Investigation, we get the data for hate crimes from 2008 to 2017.
From Gallup special report, we get the data for LGBT people by states in 2016.
From the library of congress civil war desk, we get the data for the 15 slave states before and during the american civil war from 1789 to 1861.

Additional variables explanation

After a glimpse of the seven variables provided, we found out that the cultural factor is basically not included in the regression. However, there always exists a correlation between cultural factors and realistic factors, so that it is so hard to get rid of multicollinearity. Also, culture is hard to observe and measure. We chose if_slave as a cultural dummy variable. It has several advantages: Firstly, civil war is an historical event, it

has nothing to do with a lot of variables in nowadays. Secondly, we have confidence to presume that whether or not it was a slave state have a lot to do with the state's cultural background. For example, the state of Geogia, which was a slave state, is still one of the most racist states in the United States, according to World Population Review. Also, the groups of LGBT people might also be an unresting factor that may trigger hate crimes, for the reason that some conservative people may still can not accept this group of people.

Potiential bias

Comprehensiveness

The data we collected from FBI may be biased because this data was submitted voluntarily by law enforcement agencies of each state, so a large number of crimes may not be reported. Moreover, this rogram collects data on only prosecutable hate crimes, which make up a fraction of hate incidents (which includes non-prosecutable offenses, such as circulation of white nationalist recruitment materials on college campuses).

Awareness bias

Heightened news coverage of hate incidents may have encouraged people to report incidents that they would not have otherwise reported. From the other hand, state agency may also hide a lot of hate crimes after they come to realize that more cases of hate crimes may do harm to the state's reputation, thus may cause repulsion among those people who want to invest in this state.

Definition bias

Hate crime is a quite ambiguous and obscure terminology to a lot of people, even to those who work in a law enforcement agency. So different people may have various understandings to it, and it may also varies by the different cultural background by states. SO quite a lot of officers may have different judgements on a same case.

Regression

Procedure

The first regression is the original one including seven variables provided. After adding the if_slave variable, we did a second regression. And the third regression includes all the nine independent variables.

Regression outcome

Table 1: Regression

	Dependent variable:		
	crimes_per		
	(1)	(2)	(3)
median_income	0.0004***	0.0004***	0.0002***
	(0.0001)	(0.0001)	(0.0001)
unemployed_per	−17.688	−35.236	−50.960
	(49.888)	(48.781)	(33.728)
hs_per	−38.997**	−33.833**	−20.984*
	(15.589)	(15.217)	(10.665)
white_poor_per	111.593**	119.582***	76.519**
	(44.232)	(42.768)	(30.173)
non_white_per	14.255	22.715**	13.088*
	(9.843)	(10.321)	(7.259)
non_citizen_per	−99.344***	−100.241***	−66.242***
	(24.835)	(23.919)	(17.241)
gini	118.362***	114.237***	71.517***
	(28.791)	(27.796)	(20.177)
if_slave		−2.411**	−0.535
		(1.164)	(0.849)
lgbt			217.197***
			(31.965)
Constant	−70.500***	−68.988***	−47.579***
	(16.731)	(16.128)	(11.562)
Observations	50	50	50
R^2	0.539	0.583	0.806
Adjusted R^2	0.463	0.502	0.763
Residual Std. Error	2.980 (df = 42)	2.869 (df = 41)	1.979 (df = 40)
F Statistic	7.024*** (df = 7; 42)	7.165*** (df = 8; 41)	18.515*** (df = 9; 40)
Note:			*p<0.1; **p<0.05; ***p<0.01