Hate Crimes in the United States

By: Burke Croft, Su Karaca, Paige McKindra

A. Introduction/Motivations:

In this report, we wanted to investigate more about the correlations related to hate crime rates across the United States. Specifically, we wanted to investigate how hate crimes were correlated with (1) Share of population in metro areas, (2) share of population in poverty, (3) gini index, and (4) share of voters who voted for Trump. We also wanted to investigate the influence of those 4 causal variables on the massive increase in hate crime frequency across the United States (between 2014 and 2020, instances of hate crimes in the United States increased by 84%)(Figure 1).

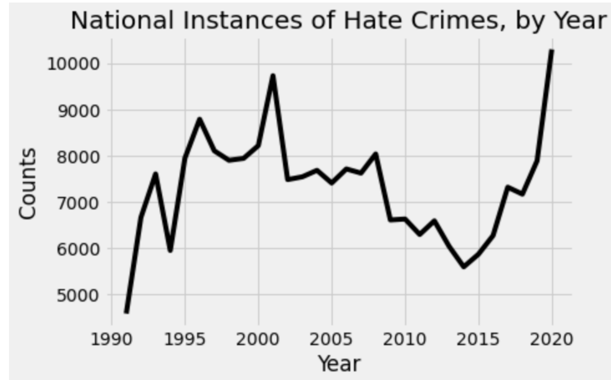

Figure 1

B. Methods:

We began by running a multivariable linear regression model of the 4 causal variables, minimizing mean standard error for FBI-sourced average hate crimes per 100k population over 2010-2015. We decided not to use the SPLC-sourced average hate crimes per 100k population due to the missing states and lack of consistency between the SPLC and the FBI-sourced data. For example in New York, the FBI reported 11% more hate crimes than the SPLC reported. We then measured recent change by running the same multivariable linear regression model of the 4 causal variables, minimizing mean standard error, but this time for 2020 rates of hate crimes per 100k population subtracted from 2020 rates of hate crimes per 100k population. Additionally, we investigated the associations between the 4 casual variables and FBI-sourced average hate crimes per 100k population over 2010-2015. We also examined the associations between the 4 casual variables and each other to check for multicollinearity.

The causal variables, as well as the 2010-2015 average, were sourced from Reference 1, a GitHub data source. For the creation of the 2014-2020 increase in rates of hate crimes, we had to pull data directly from the FBI hate crime database and census populations to produce state-by-state hate crime rates increase.

C. Results:

Figure 2. Linear coefficients for modeling hate crime rates

	Average, 2010-2015	Increase, 2014 vs. 2020
Share of Voters Who Voted Trump	-0.703	-0.601
Share of Population in Metro Areas	-0.499	-0.229
Gini Index	0.656	0.047
Share of White Poverty	-0.263	0.045

The results yielded important conclusions, including that, both for early-2010's static rates and for late-2010's increases in rates, the share of voters that voted for Trump was the most significant signifier of rates of hate crimes, *in a negative direction*. As the states became more Trumpian, their rates of hate crimes, in general, got lower. This might be due to the formation of large in-groups with the identification of Trump supporters. As an example, in the cases where the Trump supporters were the out-group, the comparatively small number of reported anti-Trump cases might be due to the fact that people who have been subjects to these incidents are less likely to report the cases to the sources and not because there are actually small number of anti-Trump hate crime cases

(Majumder). This supports our intuitive reasoning that the political inclination of the sources might be more welcoming towards the minority populations and Democrats.

For the early-2010's static hate crime rates, the gini index was the second strongest signifier, *in a positive direction*. However, for late-2010's static rates the second strongest signifier of rates of hate crimes was the share of the population in metro areas, *in a negative direction*. We could not find any research to support these two findings however, we speculate that this may be due to data collection issues discussed in the conclusion. For example, the population collected may have changed significantly between the two datasets.

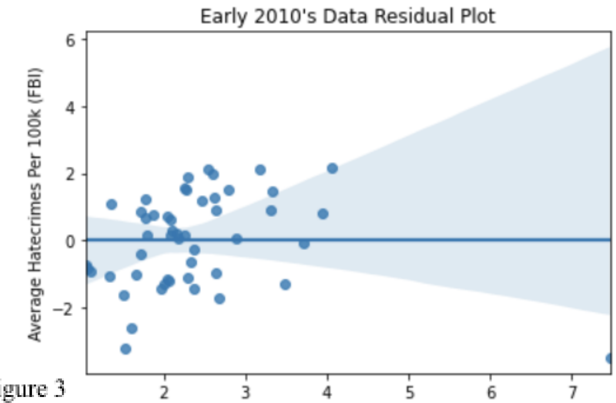

Figure 3

There were no trends found in the residual plot of the early-2010's static rates revealing the regression is a good fit for the data (Figure 3).

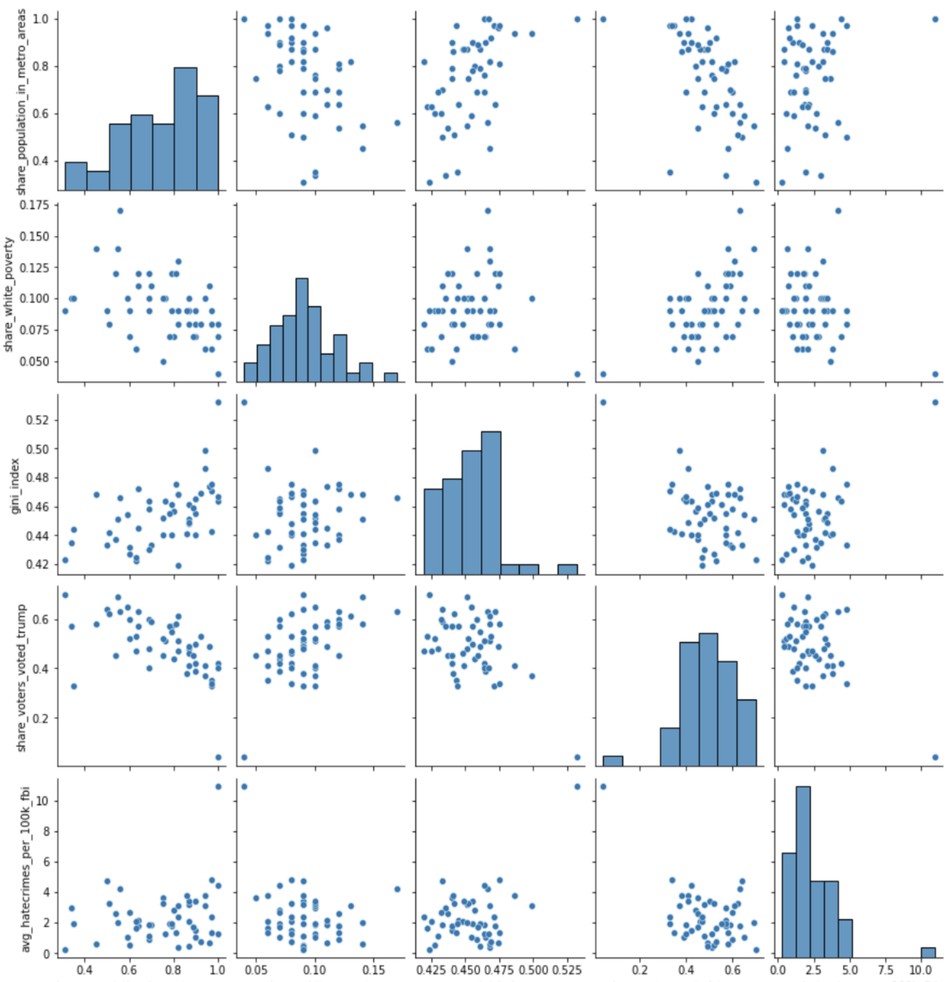

Figure 4

Running a pair plot (Figure 4) we first analyzed the associations between FBI-sourced average hate crimes per 100k population and the 4 casual variables. In the plot, the share of the population in metro areas had a weak or slightly positive association with FBI-sourced average hate crimes per 100k population. The share of the white population in poverty had no association with FBI-sourced average hate crimes per 100k population. The Gini index had a positive association with FBI-sourced average hate crimes per 100k. Lastly, the share of voters who voted for Trump had a negative association with FBI-sourced average hate crimes per 100k. There were associations between the 4 casual variables. For example, gini index and share of voters who voted Trump had a negative association. However there were no concerns for multicollinearity due to no strong associations were found between the 4 casual variables.

D. Conclusion/Discussion/Challenges:

a. Conclusion and Discussion

In Jendryke and McClure's "Mapping Crime – Hate Crimes and Hate Groups in the USA: A Spatial Analysis With Gridded Data" titled paper, it is stated that there are "at least six possible explanations for hate crime: (a) psychological, (b) social-psychological, (c) historical-cultural, (d) sociological, (e) economic, and (f) political," (Jendryke). In their spatial analysis, supporting our results about metro areas, the importance of the location of the hate crime stands out.

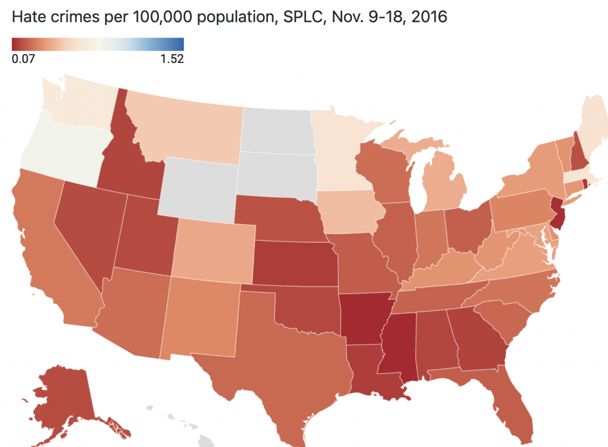

Figure 5

They explain this spatial relationship with the proximity of the hate crime locations to those of the hate groups. They state that "hate crimes and hate groups are co-located in certain parts of the US, predominantly in the South, but hate crime rates are generally high in more dense urban areas," (Jendryke). This statement can also be seen in our choropleth maps of the FBI and SPLC data (Figure 5 & 6 [via Datawrapper]). Moreover, as our data analysis shows they also consider factors such as demographic change and economic stress as factors that motivate hate crimes to a greater extent than hate groups (Jendryke). In the same report, the authors pay great attention to the incompleteness of both governmental and NGO-based sources' data, such as SPLC's. They warn that "even though [SPLC] do extensive data collection (Day, 2012), there is no indication whether the data on the website are complete or not," and share the same concerns with us, as we'll explain in the Data Collection section, about the divergent use of the hate crime definitions (Jendyke).

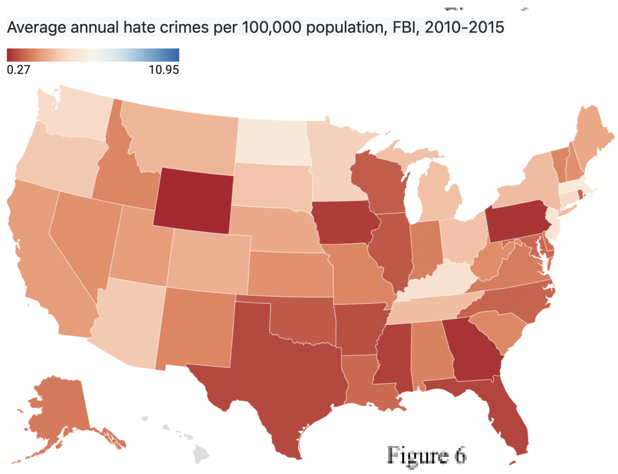

Figure 6

In conclusion, the strongest correlations with hate crimes over 2010-2015 were the share of voters who voted for Trump, *in a negative direction*, and gini index, *in a positive direction*. The strongest correlation with hate crime increase from 2014-2020 was Trump vote per share, *in a negative direction*, and share of white population in poverty, *in a positive direction*. The lack of trends in the residual plot revealed that the regression was a good fit for the data. However, due to data collection methods, it was difficult to compare data collected from the FBI versus SPLC. Additionally, it was challenging to compare data collected in the early 2010's versus later 2010's due to data originally being from two different sources (GitHub and FBI hate crime database).

b. Challenges: Data Collection Methods, Validity, and Ethics

The data used in our project groups the 'hate crimes per 100,000 population' values into categories of two sources, one provided by the SPLC (Southern Poverty Law Center), one provided by the FBI (Federal Bureau of Investigation). Because the two values of rates of hate crimes differ considerably in many states, we decided to investigate and conduct outside research to answer questions like 'why is there such a difference?', 'why certain states have lower hate crime than others?', and 'is there a contribution of the data ethics and collection methods to this divergence?'

One of the hate crime sources in the data, the non-profit organization SPLC, tracks U.S. based groups that enact or intend hateful or extremist crimes. More than 1600 active groups they track across the U.S. range from Ku Klux Klan and neo-Nazis to antigovernment militias and Christian Identity supporters (SPLC). The data presented by the SPLC includes hate categories of "Anti-Immigrant Incidents," "Anti-Black Incidents," "Anti-Muslim Incidents," "Anti-LGBT Incidents," "Anti-Woman Incidents," "Anti-Semitism," "White Nationalism," "Anti-Trump" (Miller). The hate crime incidents reported by the SPLC have two sources of submission: individual submissions made with the tag #ReportHate to their website and media accounts (Miller). To control the types of submissions made to these platforms, the SPLC does not count the cyber cases and limits their reported data to events happening in-person. About the verification process, the SPLC states that they "have excluded incidents that authorities have determined to be hoaxes; however, it was not possible to confirm the veracity of all reports," (Miller). Even though this reporting system has passed through a deliberative elimination process, the casualty of the data collection method introduces risks of over-reporting in either case.

Moreover, concerning the second hate crime source in the data, "crimes reported to the FBI involve those motivated by biases based on race, gender, gender identity, religion, disability, sexual orientation, and ethnicity," (FBI). However, because each and every state has different definitions of hate crime, "To ensure data are uniformly reported, the FBI provides contributing law enforcement agencies with user manuals that explain how to classify and score offenses using standardized crime offense definitions. Acknowledging that offense definitions may vary from state to state, the FBI cautions agencies to report not according to local or state statutes, but according to those guidelines," (FBI). Compared to the SPLC's methods, the FBI presents a more structured approach to their data collection strategies. However, this structure might have the consequence of data loss if the reported hate crime does not fit the manual's definition of a hate crime. So, even though the FBI database collects reports from "more than 18000 city, university and college, county, state, tribal, and federal law enforcement agencies eligible to voluntarily report criminal data," (FBI) due to the strict cleaning processes, there may be an underrepresentation of the actual numbers of hate crimes.

The SPLC reports that according to The Bureau of Justice Statistics, approximately "two-thirds of hate crimes go unreported to the police," (Miller). This problem becomes more pronounced when the cases under investigation are events that do not carry the importance of "criminal violation" and/or reported to a less acknowledged repository (Miller). The very same problem was under investigation in the early 2000s, when misclassification and underreporting issues of the law enforcement agencies that are connected to NYPD began to increase in number. The NYPD went under an audit as it was "unable to confirm that all reported bias incidents are properly captured, recorded, and reported," (McDevitt). These questions resulted in the shattered image of the law enforcement agencies' data keeping credibility. On the flipside, the problem with the SPLC's data keeping method was the "decentralized and uncoordinated approach," which resulted in "each group collect[ing] data of different types," (McDevitt). There was no clear definition of a hate crime or the identity of the victim (who was the victim –a group or an individual?)

E. REFERENCES

1DATA.Fivethirtyeight. "Data/Hate-Crimes at Master · Fivethirtyeight/Data." *GitHub*, github.com/fivethirtyeight/data/tree/master/hate-crimes.
2DATA.FBI. "FBI Crime Data Explorer." Crime Data Explorer, crime-data-explorer.fr.cloud.gov/pages/explorer/crime/hate-crime.
3.FBI. "About the UCR Program." FBI, 15 Sept. 2018, www.fbi.gov/file-repository/ucr/about-the-ucr-program.pdf/view.
4.Jendryke, Michael, and Stephen C. McClure. "Mapping Crime – Hate Crimes and Hate Groups in the USA: A Spatial Analysis with Gridded Data." Applied Geography, vol. 111, 2019, p. 102072., doi:10.1016/j.apgeog.2019.102072.
5.Majumder, Maimuna. "Higher Rates of Hate Crimes Are Tied to Income Inequality." FiveThirtyEight, FiveThirtyEight, 23 Jan. 2017, fivethirtyeight.com/features/higher-rates-of-hate-crimes-are-tied-to-income-inequality/.
6.McDevitt, Jack, and Janice A Iwama. "The Handbook of Measurement Issues in Criminology and Criminal Justice." Google Books, Google.
7.Miller, Cassie, and Alexandra Werner-Winslow. "Ten Days after: Harassment and Intimidation in the Aftermath of the Election." Southern Poverty Law Center, 29 Nov. 2016, www.splcenter.org/20161129/ten-days-after-harassment-and-intimidation-aftermath-election.
8.SPLC. "Fighting Hate." Southern Poverty Law Center, www.splcenter.org/fighting-hate.
9. U.S. Census Bureau, Population Division, Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2019; April 1 2000 to July 1 2009; April 1 1990 to July 1 1999
10. FBI, Hate Crime Statistics (https://www.fbi.gov/services/cjis/ucr/hate-crime