Section 2: Week 5: Interpreting Results

Nate Bachmeier

TIM-7101: Statistics for Technology Leaders

August 2, 2020

Northcentral University

# Interpreting Results

NCU-Cares (NCU-C) is a politically neutral non-profit that seeks to improve the work through targeted lobbying efforts. The death of George Floyd has risen the debate of police violence and reform to the national stage (Crary & Morrison, 2020). While the topic rests on American’s hearts and minds, it has also become highly partisan with many efforts to undermine the conversation (McCaskill, 2020). After collecting and assessing information from the Washington Post, the organization needs to frame those results into actionable recommendations and enumerate any limitations. Limitations always exist in research due to insufficient time and resources to consider all aspects of a problem (García-Pérez, 2012). Through clarification of these issues, researchers can reduce the impression of being bias or misleading and guide further conversations.

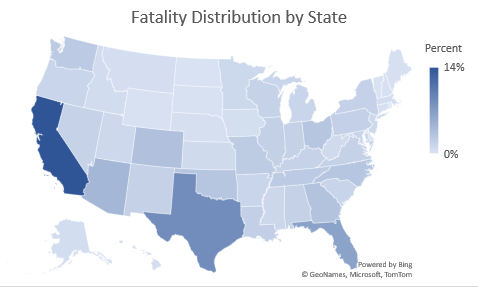
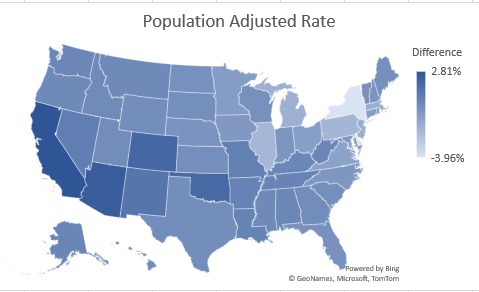
# What are the observations

## Expected Outcome

The Washington Post’s collects demographic, location, and contextual information on victims of police violence from January 2015 through the present. Researchers and media sources use this data to quantify the racial basis of law enforcement. Their observations become statistical statements such as “unarmed Blacks are twice as likely as Whites to be the victim (Nix, Campbell, Byers, & Alpert, 2017).” Another frequently cited metric that “Black people were 24% of those killed despite being only 13% of the population (KBP, 2020).” Over 200 publications that reference Nix et al.’s paper in Nature magazine and Google returns 7.7 million results for the second quote. When NCU-C set out to assess the problem, the initial expectation was to find racial inequality and bias decisions against people of color. While both conclusions are true, that does not appear to be the entire story about police brutality in America.

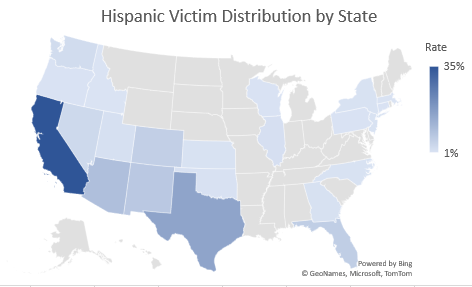
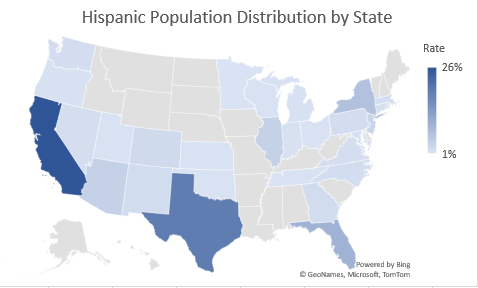
## Actual Outcome

Figure 1: Fatalities by State (2015-2019)

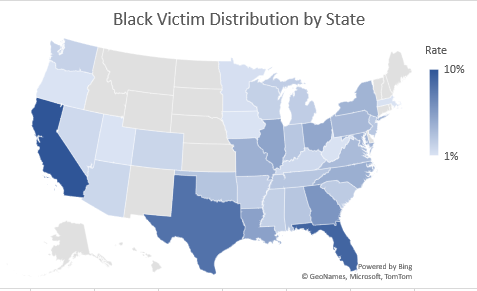
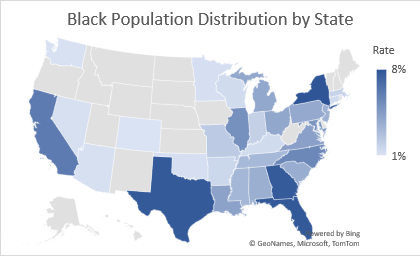
NCU-C’s investigations show that the method of sub-grouping and classification has significantly more impact on the results than any arbitrary feature. For instance, of the 4937 records, California and Texas own 14% and 9% of incidents, respectively (see Figure 1). There might be specific challenges within those states; however, they are also home to 12% and 9% of the national population (Census Bureau, 2019). Assuming all other things are equal, the actual number of fatalities less expected fatalities based on population shows the rates are relatively stable across the country.

Figure 2: Hispanic Victims by State

Next, an analysis of the total number of Hispanic victims in each state shows a strong correlation with the group’s natural population distribution (see Figure 2). This perspective highlights some of the challenges associated with treating the country as a homogenous cluster. Instead, decomposing America into regions allows for a more accurate assessment of the relevant population counts to include. For example, since at least 2015, no police officer has killed a Hispanic person in Montana (Washington Post, 2020). However, this group also only represents 38,000 (less than 4%) of that state’s inhabitants (Census Bureau, 2019). Then consider that Texas has over 11 million (39.4%) residents of Hispanic and Latino descent. Police violence in this state has killed 432 people during this same period, of which 143 (33%) were Hispanic.

Figure 3: Black Victims by State

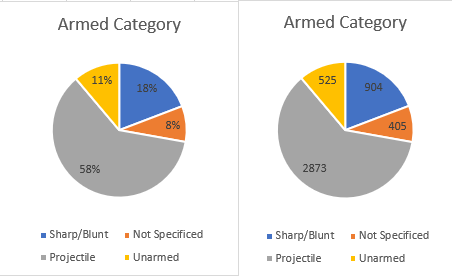
 

An extrapolation of comparable ratios exists in many other locations and ethnic groups. For example, approximately 39 million Black citizens live across the United States, of which 1277 have been killed by police violence since 2015. By plotting the distribution of these groups in terms of population per state, results in similar charts (see Figure 3). These pivots suggest that the deaths are not racially motivated, and instead, a function of the locale-specific population make-up. If that was not the case, then more pronounced outliers should exist with substantially higher victim rates relative to the group’s population. This outcome also hints that some other motivator outside of the *race* is likely to exist.

## Potential Motivators

There are fifty-three million people who have an interaction with the police each year (BJS, 2015). Annually, 167 (+/- 6) officers and 66 (+/- 16) unarmed civilians die during these altercations (NLEOMF, 2019). These situations have to introduce a wide range of emotions and hostility, which creates challenges while balancing the safety of officers and civilians. NCU-C should invest additional resources into exploring solutions to reduce this friction. For instance, of the incidents that result in death, 76% (3853 of 4937) involve the suspect having a weapon (see Figure 4). Training and procedures could exist to address these scenarios in a manner that improves the probability of a peaceful resolution. While it does not address all of the concerns, this is a significant source of police brutality. Specifically, the unarmed and unspecified groups need further consideration to reduce the loss of life for these subgroups. However, it might be impossible to eliminate due to this group representing 0.00002% (119 of 53 million) interactions per year.

Figure 4: Suspect Armed Category 2015-2019 Total



# Research Limitations and Challenges

The four major threats to research projects are internal, external, statistical conclusion, and construct validity (Parker, 1993). While efforts to minimize these risks do exist, the time and resources of this project are relatively finite.

## Internal Threats

An internal threat exists when the researcher does not accurately represent the results. This scenario could exist from biases during the categorization and groupings of the victims. For instance, the Washington Post data set uses a free-form text field to record the weapon and threat level. The analysis uses a few general buckets versus other researchers propose using more fine-grained options (Nix, Campbell, Byers, & Alpert, 2017). Minor forms of selection bias might also exist due to the filtration rules of the data set. Expressly, the analysis only includes records that provide the racial demographic and location of properties.

## External Threats

An external threat comes from a variable that is outside of the researcher’s control. The government does not require law enforcement agencies to report incidents that result in police brutality. Since official sources do not exist, researchers must rely on open-source data sets like the ones provided by the Washington Post. The Post uses news and social media reports, which could be both erroneous and lossy. There are also risks that the manual entry process could have inaccurate values for a record in the table (e.g., wrong *race*). Another class of risks comes from the data set being immature and starting in 2015. Ideally, having more longitudinal data to understand trends or alternative sources for cross-validation would improve the validity of results.

## Statistical Conclusion

Figure 5: Victims by Race



Invalid statistical conclusions arise from not having sufficient samples or encountering too many uncontrolled parameters. The filtered data set only contains 4937 records, which does not provide sufficient evidence for some pivots (see Figure 5). For example, assessing Asian and Native American victims independently likely result in overfitting. Due to time constraints, the evaluation of only a subset of features took place, and this could have missed an important or confounding variable (e.g., age group).

## Construct Validity

Threats to the construction of the experiment occur when controls do not protect against information leaks between tests or controls between variables. These risks might exist due to the analysis evolving with the research project. While a general outline and strategy exist, the budget to procure sufficient causes a more relaxed set of requirements than during the onset. It is also possible that initial observations encouraged exploration of specific portions of the dataset because it more easily aligns with the topic. Instead, a more thorough effort could exist that examines other pivots and asks other questions of the information.

# References

BJS. (2015, October). *Contacts between police and the public*. Retrieved from Bureau of Justice Statistics: https://www.bjs.gov/content/pub/pdf/cpp15\_sum.pdf

Census Bureau. (2019). *Quick facts*. Retrieved from United States Census Bureau: https://www.census.gov/quickfacts/fact/table/US/PST045219

Crary, D., & Morrison, A. (2020, June 11). *Black Lives Matter goes mainstream after Floyd's death*. Retrieved from AP News: https://apnews.com/347ceac3ea08978358c8c05a0d9ec37c

García-Pérez, M. A. (2012). Statistical conclusion validity. *Frontiers in Psychology, 3*. doi:https://doi.org/10.3389/fpsyg.2012.00325

KBP. (2020, August 11). *Police shootings database*. Retrieved from Killed By Police: https://killedbypolice.net/

McCaskill, N. (2020, July 16). *Trump has tried to make 'Black Lives Matter' a partisan issue. A federal agency says it is not.* Retrieved from Politico: https://www.politico.com/news/2020/07/16/trump-special-counsel-black-lives-matter-366177

Nix, J., Campbell, B., Byers, E., & Alpert, G. (2017, February 8). A bird's eye view of civilians killed by police in 2015. *Criminology and public policy, 16*(1), 309-340. doi:https://doi-org.proxy1.ncu.edu/10.1111/1745-9133.12269

NLEOMF. (2019, April 10). *A year-by-Year breakdown of law enforcement deaths throughout U.S. history*. Retrieved from National Law Enforcement Officers Memorial Fund: https://nleomf.org/facts-figures/officer-deaths-by-year

Parker, R. (1993). Threats to the validity of research. *Rehabilitation Counseling Bulletin, 36*(3), 130-138. Retrieved from https://search-ebscohost-com.proxy1.ncu.edu/login.aspx?direct=true&db=eric&AN=EJ458938&site=eds-live

Washington Post. (2020, August 1). *Data police shootings*. Retrieved from GitHub: https://github.com/washingtonpost/data-police-shootings/commit/2ea87a4ed725b164be1489878e20827fb5944d1b