# Lab 3 Peer Feedback

Reviewers: Group 6 (Namburi, Pouchepanadin, Reyes)
Reviewing: Zach Merritt

## Introduction (Section I)

Could possibly add more context (e.g. who you are, who is the candidate) or clarify the goal is to reduce crime.

## Initial EDA (Sections II, III, IV, V, VI)

**Section II**
No issues with this section

**Section III**
No issues with cleanup of NA values, but were there any other checks for other, non-NA anomalous values?

Could consider analyzing outliers in the dataset or variables of interest. These findings in the exploratory data analysis could help analyze the impact of the outliers in the models. Cook's distance graphs could be leveraged to support the analysis.

**Section IV**
Great approach identifying the relationship with FIPS codes. Could possibly talk more about how you confirmed this relationship just to better reassure readers that the mapping isn't coincidence (e.g. specific pattern of every other odd, data is a subset, choropleth maps & geographic indicators align, etc).

The use of choropleth maps. The maps are visually engaging and makes it clear that the missing dummy variable is the East region.

Formatting could be improved for this section to save wasted space. Page 9 only contains a single line.

There appears to be a county that is flagged as both central and west. The county sticks out especially on the west choropleth as it is isolated from the rest of the group.

Since report length is an issue, may want to consider the usefulness of the maps when looking at the density of the independent variables.

**Section V**
Very clear and systematic approach, great use of skewness and kurtosis

Extensive use of visuals eat up a lot of pages. Since page length is an issue and this content primarily addresses data transformation, it would be best to trim this down to allow more focus on models.

**Section VI**
Could potentially explain the purpose of analyzing the correlations across the independent variables in terms of exploring presence of multicollinearity, which could affect the model building process.

## Model Building Process (Section VII)

Although mentioned in the comments of the code, should point out in the text that some initial cleanup is being performed prior to constructing the models.

May be good to support how minority populations can serve as a proxy for income by citing a study or statistic.

Report makes strong claims about racism in America. While the data does show a significant correlation, it would still be good to be a little more nuanced (e.g. instead of "proof of racism", possibly "supporting evidence of racism" or “possible indicator of such racism”)

Could explain more as to why highly collinear variables were removed, just to show explicit knowledge of course topics

It's interesting that the AIC for model 2 is lower than model 1. How could this be explained?

Could consider analyzing adjusted r squared for the models.

The idea of building a causal model with the regression equation seems to be skimmed over in the model specification phase.

## Regression Table (Section VII)

Formatting of regression table could be improved, it seems to cut text in half.

Could include star cutoffs to indicate statistical significance and adjusted r squared.

## Omitted Variables Discussion (Section VII)

Expected more explanation on the omitted variable side. The requirements were to discuss 5 to 10 omitted variables and the direction of the bias.

## Conclusion (Section VII)

The final conclusion is light and could do more to address the bigger picture. Some questions that could be explored:

- Is this single study enough to act on or do you recommend additional analysis?
- If another study were to be done, what are some changes you would recommend?
- What are the logistics of the suggested policy?
- How hard would it be to enact such policy?
- What are the shortcomings of the policy?
- Is there still value in non-causal variables within policy making?

Given the significant focus on density and crime rate, you could suggest steps/policies to reduce density in areas with a high crime rate, e.g. create job opportunities in other counties, provide incentives for industries so jobs/people could move.

## General/Other Feedback

The report is 33 pages long, significantly longer than the page limit indicated on the lab instructions. The report may need to be trimmed down, especially since there are still elements to be added for the final report.

Section VII seems to address too many aspects of the entire lab. The omitted variables discussion and conclusion could be separated into their own sections.

Overall, liked the way and the code Zach used to attack the problem. There could have been more attention paid to writing out or explaining the different sections in the report