1. Introduction. As you understand it, what is the motivation for this team's report? Does the introduction as written make the motivation easy to understand? Is the analysis well-motivated? Note that we're not necessarily expecting a long introduction. Even a single paragraph is probably enough for most reports.

- Having demographics as part of the research question might not be best. Demographics are generally in the control of local governments. The sentence before does a better job with the statement about "identify determinants of crime to help develop crime control policies." A better research question might be "What are the causes and determinants of the crime rate in North Carolina"

- Measuring the effect of punishing crime as a mechanism for policy makers to control crime is something that would be a necessary action but is unlikely to identify preventative action that could be taken to address the motivation and environment variables that lead to someone performing a crime, if such opportunities exist. Large bodies of research indicate crime is highly correlated with youthfulness and early involvement in crime is a good predictor of subsequent involvement. Inequality, disrupted families, poverty, criminal activity and inadequate socialization are some of the root causes and would need to be addressed by policy makers in North Carolina.

2. The Initial EDA. Is the EDA presented in a systematic and transparent way? Did the team notice any anomalous values? Is there a sufficient justification for any datapoints that are removed? Did the report note any coding features that affect the meaning of variables (e.g. top-coding or bottom-coding)? Can you identify anything the team could do to improve its understanding or treatment of the data?

- Overall the EDA is great, my only concern is about the justification to not use the log of a variable. "we will not transform the Crime Rate variable, given there are 79 observations (after EDA/data transformations) and it meets the 'n=30' requirement of the Central Limit Theorem." The choice to use or not use a transformed variable has nothing to do with the CLM, rather it's about the underlying relationships in your (theoretical) population model). If in your population model you expect a heavy right skew to the crime rate, then you should use a transformed variable, as it better captures the underlying relationship.

- I concur with the other reviewer, I found the EDA to be very well laid out, concise, clear and logical in its approach to the task and addressing any issues found.

  I also thought the process of comparing various variables to create a univariate and bivariate analysis helpful in understanding the data and the thought process the team went through to create its analysis.

  The correlation analysis was interesting and the team used this to reject their hypothesis. It did not offer any alternative based on their findings which would of helped me as a

reader understand what the group subsequently gleaned from this analysis and how their point of view may have changed based on what they found.

3. The Model Building Process. Overall, is each step in the model building process supported by EDA? Is the outcome variable (or variables) appropriate? Did the team consider available variable transformations and select them with an eye towards model plausibility and interoperability? Are transformations used to expose linear relationships in scatterplots? Is there enough explanation in the text to understand the meaning of each visualization?

- Overall everything is explained very clearly and I liked that the team laid out what they expected to happen as a result of the model building exercise and the outcome. My only suggestion would be to look more at transformations. Given all the heavy right skews, using log or other box-cox transformations would likely lead better correlations/models.

- More discussion of why there is a positive relationship between crime and cops would be helpful; it seems that this is clear example of OVB.

4. The Regression Table. Are the model specifications properly chosen to outline the boundary of reasonable choices? Is it easy to find key coefficients in the regression table? Does the text include a discussion of practical significance for key effects?

- The text includes a description of the analysis of the data within the regression table and the results from the AIC test and the decisions made.

5. The Omitted Variables Discussion. Did the report miss any important sources of omitted variable bias? For each omitted variable, is there a complete discussion of the direction of bias? Are the estimated directions of bias correct? Does the team consider possible proxy variables, and if so do you find these choices plausible? Is the discussion of omitted variables linked back to the presentation of main results? In other words, does the team adequately re-evaluate their estimated effects in light of the sources of bias?

- I think Crime Type is the largest omitted variable they didn't talk about. For some crimes, such as shoplifting the probability of arrest/conviction would likely have a huge impact on how likely an individual is to commit the crime. On the other hand, for crimes of passion like Voluntary manslaughter, an individual isn't thinking about the results of their actions and their decision likely wouldn't be as impacted by the chance of getting caught.

6. Conclusion. Does the conclusion address the big-picture concerns that would be at the center of a political campaign? Does it raise interesting points beyond numerical estimates? Does it place relevant context around the results?

- The conclusion lacks discussion what I see as one of the most important factors in all models - the high coefficient for number of police (~11). Given this result there are two possible explanations: 1, police actually cause crime and the number of police should be reduced (nonsensical), or 2, there are huge omitted variables skewing the results. Given the data presented, I don't understand what supports the recommendation to "optimise over staffed police offices in counties with lower crime rates and reallocate the officers to counties that appear under-staffed based on our model." This recommendation seems to be to based on the prior assumption that police help reduce the crime rate, which isn't support by the results presented.

7. Throughout the report, do you find any errors, faulty logic, unclear or unpersuasive writing, or other elements that leave you less convinced by the conclusions?

- The report is well written and concisely assembled explaining all of the steps the team took, I found the document easy to read and follow along with the team as they discussed their analysis and findings.

  The only comment which is highlighted in point 6 is such that the team would take a disciplinary approach to dealing with crime vs. looking to address / research the root causes of crime which may result in a more positive and longer term strategy to tackle crime, this would include research into items such as wage, education, employment and housing by location and other key variables.