Team 08

The current analysis examined the presence of racial disparity in the SQF program by leveraging on weapon recovery rate. The use of weapon recovery rate is innovative and well justified, and the results surprisingly suggest that there existed no racial discrepancies in weapon recovery rate upon adjusting for other factors. The report is overall well-written. Some sections, however, can use some more organization and expansion to improve its clarity.

Introduction:

While Section 1.1 contains enough supporting the current approach, the flow is a bit difficult to follow at first glance. The authors started with hit rates, shifted to CPW crimes, and weapon discovery rate, claiming weapon discovery rate is a more subjective measure than "hit rates" in other types of crimes. It was unclear what hit rates refer to in this section, as it seems to be used interchangeable for arrest rate and weapon recovery rate? In my opinion, it might give the readers a clearer picture to start justifying focusing on CPW data only (maybe back up with some literature). Then bring in weapon recovery rate as a measure for racial disparity using CPW data. The general introduction (i.e. paragraphs under 1 Introduction) can also be catered more towards CPW.

Spatial effects seemed to have come out of nowhere; it was mentioned a few times under Sections 1 and 1.1 before being explored in Section 1.4. Is there any literature that can be cited and discussed that support the control of spatial correlation? Meanwhile, Section 1.4 contained some nicely made graphs. However, I do think that it is worthwhile connecting the findings from the exploratory data analysis to the narrative (instead of briefly explaining them only in the figure captions). Overall, I think more justification should be provided for analyzing/controlling for spatial dependency.

There are also a few minor issues worth noting. First, the disclaimer under 1.1 is nice to include but should probably go in the discussion when talking about the limitation of the analysis. Second, weapon recovery and discovery seemed to be used interchangeable (throughout the entire report). It is probably nice to be consistent in wording to avoid confusing the reader.

Methods

There are a few questions I have regarding the authors' choice of methodology. First, under Section 2, the authors mentioned that "there is also a justification for the spatial lag model," on top of the spatial error model. Why didn't the authors considered using a SARMA model, which can account for both error correlation and lag? Overall, I think there should be more theory /reason-based justification for why a spatial-error model is chosen over both spatial lag and SARMA models.

Second, what are some justifications/rationales for the choice of predictors included in the analysis? For example, why is precinct population included? In addition, why did the authors decide to combine proportions of Blacks and Hispanics among the CPW sample into one group, but break those two groups into separate categories when it comes to the precinct population? Why was `Majority White` included, when Black and Hispanics population at the precinct level has already been controlled for? Doesn't the Black and Hispanic population proportion already provide sufficient information regarding whether Whites are the majority in that precinct? The inclusion of proportion of CPW suspects who were minorities and males makes sense but should probably be explained.

From the correlation matrix, it does appear that proportion of Blacks at the precinct level is pretty significantly correlated with percent of Blacks and Hispanics stopped by SQF, which could potentially result in multicollinearity issues. I wonder if making Black and Hispanic precinct-level population into categories might help address this. For example, the authors can consider condensing the numerical values into categories such as "less than 10%" and "more than 10% & less than 40%."

There are also a few suggestions for organization/minor concerns:

a) the presentation of information in the Methods section can also be reordered to improve clarity and flow. Right now, the paragraphs under Section 2 seems to be a bit disorganized. The authors started claiming that they fit a spatial error, ran tests to find that both spatial error and lag models are possible, and then decided to choose the spatial error model. I think a better way to organize this may be to start with justification for controlling for spatial dependency and reasons to choose a spatial

error model. Then conduct the test to support this choice. It is better to present the choice as research/theory-driven than data-driven.

b) The authors might want to specify which predictors are the most important to answering the research questions. Right now, there are four race-related variables: prop of CPW suspects who were Blacks or Hispanics, prop of Hispanic residents, prop of White residents, and Majority white. Which one of these predictors, if found significant, is the most indicative of racial disparity?
c) For the model specification under Section 2, I think the spatial error term (lambda*W*u) does not match the dimension of the outcome (weapon discovery rate for ith precinct), as W is an nxn matrix. To correct this, indices should probably be added to this term. Similarly, the error term epsilon should also have an "i" subscript.
d) Lastly, the model assumptions are mashed together under Section 2.1 without proper punctuation.

Results

The results section is overall simple and straightforward. It may be worthwhile summarizing that none of the race-related predictors are significantly associated with weapon recovery rate in this section. The authors talked about the spatial error coefficient being significant, but the research goal is whether racial disparity exists in CPW stops. Mentioning briefly what the race predictors suggest could facilitate reader understanding of the results in relation to the research question.

Discussion

This section contains some nice discussion regarding limitation and future direction, but I think more implications and explanations for the findings can be included. The authors did a great job interpreting the results, but what are the implications of such results? For example, what could it possibly mean that there exists high degrees of spatial correlation among precincts? What are some potential explanations for this finding, and what is the takeaway?

In addition, the authors could consider talking about how the finding that there exists no racial disparity in CPW stops tie into the current literature. In other words, is this findings consistent with the extant literature? Overall, I think discussing the findings in the context of prior works would help the readers understand the practical significance of the results better.

Furthermore, what are some possible explanations for the findings, and does it really suggest that people from different races are stopped equally for CPW crimes? For example, could it be that the correlation between the race predictors caused none of them to turn out as significant?

Appendix

The graphs contained in this section are beautiful and organized, and here are some minor issues I noticed. First, from the diagnostic plots, there seems to be an outlier. Could this outlier be an influential point? Second, Figure 3 is very informative, but the variable names

are a bit difficult to read. Lastly, there seems to be an extra variable name in the table under 6.4.

There unfortunately seems to be an issue with reproducibility of the code; when trying to knit the rmd, it gives an error at line 256 saying “Cannot identify variables.”

STA440 Case03 Team08 Peer Review

[redacted]

March 2021

Dear Editors,

We have reviewed the authors' paper and code. We found the paper to be well articulated and sufficient. The strengths of the paper were its introduction and discussion of the results. There were areas of improvement in formatting and the methodology section that we will discuss below. Overall, we hope this review can be beneficial for the authors in presenting their analysis.

We first begin with a discussion around the introduction. The authors' introduction was effective in describing the goals, questions, and hypothesis of the analysis. A strength of the introduction was the authors' discussion around why focusing on CPW events was a more effective way of analyzing if racial disparities exist. By discussing prior literature along with stating three reasons in support of focusing the analysis on CPW events, this helped readers who might have been more familiar with previous literature that focused on stop, search, and frisks as the response variables to see the benefits of focusing the analysis on CPW events. We also appreciated the authors' note on this analysis can not make any causal claims since this helped make it clear to the readers the type of conclusions this analysis could yield. The authors also did well in discussing the data from its source to important variables. Furthermore, the authors present two visuals, and we found both to be sufficient in motivating the authors' analysis. We do believe that the second visual has room for improvement. The authors display a red line with the text, "mean: (insert value)" in the visual; however, the text and the line overlap, making it not pleasing to read. We believe that moving the text to one side of the line can improve the readability of the visual; thus, we suggest the authors shift the text to start from the right side of the line.

Transitioning to the methodology, we found the authors' methodology to be sufficient and effective for the most part in answering the questions of the analysis. The authors did well in explaining why they chose a spatial error model over alternatives. Also, we appreciated the discussion around using a spatial lag model as a sensitivity analysis given the inconclusive test results. The authors also effectively discuss and validate the assumptions. The methodology did contain an error in how the authors denoted their spatial error model. The authors express their error model in the following format:

$$\begin{aligned} \text{weapon discovery rate}_i = \beta_0 + \beta_1 &\text{Prop Blacks and Hispanics stopped }_i + \beta_2 \text{Prop males stopped}_i && (1)\\ &+ \beta_4 \text{Prop Hispanics residents}_i + \beta_5 \text{Prop Black residents}_i && (2)\\ &+ \beta_6 \text{MajorityWhite=Yes}_i && (3)\\ &+ \beta_7 \text{Total Population}_i + \lambda \mathbf{Wu} + \epsilon && (4) \end{aligned}$$

However, this is wrong as this is adding a scalar with a matrix $\lambda \mathbf{Wu}$. We suggest the authors change the current formula notation to the following matrix notation:

$$\text{weapon discovery rate} = X\beta + \lambda \mathbf{Wu} + \epsilon \quad (5)$$

In addition to this change, the authors should also describe what one row of the X matrix would look like to show which variables were included. Another improvement to the methodology would be to describe in more depth the OLS model. The first mention of the OLS model is in the following line, "We compare the

regression coefficients with those from the spatial lag model as well as those from the OLS model." To the reader, this lack of a previous mention of this OLS model makes it unclear what the setup and role of the OLS model are. Thus, we recommend the author explain the role and denote the formula for the OLS model in the sensitivity analysis in order to clear up any confusion.

The results and interpretation of the results were effective and correct. We found the table the authors used to be professional and effective. Furthermore, we appreciated the author detailing important units such as "Precinct Population (in 10,000s)". By doing so, this made it clear to the reader about the scale of the variables.

We also found the discussion section to be sufficient in answering the research goals. The authors did well in using their sensitivity analysis to further support their results. The authors also did well in synthesizing the results from their analysis; however, the authors can improve their synthesis by providing real-world implications of these results. Currently, the authors provide no real-world discussion of what a lack of significance for their variable, proportion of CPW Suspects who were Blacks or Hispanics, means to lawmakers. We believe further discussion on the real-world implications of these results would strengthen the paper by making its results more impactful; thus, we recommend the authors discuss in section 4.1 how a politician can use their results to better make laws in the future relating to the SQF program. Furthermore, another improvement to the discussion is adding discussion around the OLS model. The authors describe the significance of the lambda parameter and lack of significance for the variables from the spatial lag model. The authors do not discuss in much depth about the OLS model despite stating in their methodology that they will compare the results of the OLS, spatial lag, and spatial error models. We believe by discussing the OLS model in more depth in the discussion section, the authors can clear up the exact role of the OLS model (is it to support results on variable significance or as a model to compare for appropriate fit) and moreover do as they described in the methodology. Thus, we recommend the authors add model interpretations and discussion of how the OLS model results impact their current results near the end of section 4.2.

Following the discussion section, the authors clearly discuss limitations, strengths, and next steps. In terms of limitations, the authors discuss how the focus on weapon discovery could have limited the analysis in terms of inference on other crimes. The authors also clearly describe how the variables that they controlled were another limitation. We appreciate how the authors in their discussion of the limitations highlight possible solutions such as to solve granularity concerns, they could use Zip codes instead of precinct instead. The strengths of the analysis are the use of the spatial error model to account for underlying spatial dependencies and the use of the spatial lag model as a sensitivity analysis. We also appreciated the authors providing examples of next steps such as focusing on drug-related stops.

We also found the paper to be professionally presented and generally free of distracting errors. We did find some distracting formatting or grammar issues that we believe the authors should address. First, the authors have unusual formatting for describing their model assumptions. The authors describe the assumptions as follows, "Linearity between response and each of the predictors Normality in residuals Constant variance in residuals Independence in outcomes of the observations". We suggest the authors change "Linearity between response and each of the predictors Normality in residuals Constant variance in residuals Independence in outcomes of the observations" to linearity between response and each of the predictors, normality in residuals, constant variance in residuals, and independence in outcomes of the observations." Another issue was the authors did not capitalize Black and Hispanic in the discussion section. We suggest the authors change "black" to "Black" and "hispanic" to "Hispanic". By implementing these two changes, the authors can remove distracting errors, leading to a more professional paper.

Furthermore, the analysis was also fairly easy to reproduce. All the necessary data files and code were located on their GitHub repository. Downloading the files and reproducing the results took about 15 minutes. Overall, there were no challenges in reproducibility.

Ultimately, we found the authors' analysis to be very interesting because of their focus on CPW events instead of arrest or frisk events. Given, how we focused on frisk, this analysis has highlighted an interesting

dimension from which to supplement our analysis. Inspired by this analysis, we could split stops into those that involved a CPW, those that involved drugs, and others. From which, we can analyze how frisk rates compare across each of these types of stops. In terms of general suggestions, we recommend the authors clarify the role of each model used in the analysis and be careful with model notation.

Case 03 Team 08 Peer Review

[redacted]

The following is a peer review of Case 03 Team 08's report. Overall, the authors produced a very impressive and well-rounded paper. Their introduction (section 1) starts off with a comprehensive literature review, where they give an overview of the policies and controversies surrounding racial disparities in SQF events. Though they highlight the importance of highlighting such racial profiling, linking frequent contact with the police with anxiety symptoms and trauma, they move onto their research analysis without justification on why investigating spatial effects at a precinct level is needed. The introduction would have better flow leading up to their hypothesis if they were to provide literature review regarding SQF incidences for particular precincts or neighborhoods that give cause to conduct spatial analysis. Their cited author Gelman, Fagan and Kiss give a thorough background on SQF events in neighborhoods where majority of residents are ethnic minorities. Lastly, in further sections the authors justify focusing on Criminal Possession of Weapons since it "bypasses the problem of hit-rate fallacy, which argues that police are biased against White people, since they have a higher arrest rate". This statement is not cited, and additionally if such was the focus of their report, then it might be useful for the reader to have additional literature review supporting this claim in the introduction section. Currently, their introduction provides a history of SQF in NY but does not tie it to connections between racial disparity and CPW.

Their next section (1.1) is labelled 'Approach', though it may be a little misleading since it also talks through their data-wrangling process. Overall, however, they provide a reasonable justification for how they chose their outcome variable, which is weapon found on suspect, as they provide further literature review and cite a similar statistical analysis that they adapted from for this case study. One note of improvement for this section would be to make it explicit in this section what their response variable is, since they define certain variables like 'Hit Rate' and 'Weapon found on suspected individual', as well as 'CPW'. The authors do note it in a later section (1.4), however stating it briefly in this section would make the paragraph easier to follow for the reader. The authors' selection of response variable is unique, and they state that 'weapon discovery is a direct indicator of whether a police officer's suspicion of CPW is justified'. This seems to be both valid but also questionable, since different states in the US have different laws for possession of weapons, and many Americans justify carrying weapons with their right to own arms as well as a means of self-protection. In addition, this would make this case-study very specific to American laws (which is fine), since in many other countries around the world, carrying any type of weapon is illegal. Lastly, what might have been useful in this section would be to define the list of weapons that would be considered criminal. Many individuals may carry pepper spray or a knife as a means of self-protection, but purchasing these items is not a violation of the law. Again, their argument of using CPW bypassing the "problem of hit-rate fallacy" is not cited.

Section 1.2 'About Data' talks about the source of their data, however lacks details about their data-wrangling process, since they mention in their introduction (section 1) that they are investigating spatial effects at the precinct level. Currently it is unclear what their independent variables are and how they were calculated. It is also not stated that such aggregation of data at

precinct level is needed in order to conduct spatial analysis. They authors have additionally not cited where they obtained the 2020 New York census data. One note is that the authors mentioned they have found an additional dataset from the NYC planning website, which includes data on the most recent police precinct boundaries, yet again, it is unclear how they merged this dataset with all others in order to have aggregated data.

Section 1.4 'EDA' begins with a choropleth map of precinct-level weapon recovery rate in NY. Though they argue that visually they observe spatial clusters, the map actually seems a little mild in illustrating that. We do not observe any high concentrations of weapon recovery rate anywhere, and all the proportions per precinct are moderate. One note for improvement would be to specifically identify the regions where we observe high clusters. The reader may not be familiar with NY, and even brief statements like "especially in regions of the Bronx, we observe some clustering" may be useful. Their second figure was on the other hand extremely impressive and helpful, where they provided a distribution of precinct-level weapon discovery rate for ethnic minorities and white people. This was especially a EDA visualization I would like to have incorporated into my own report. Given that the EDA so far then seems to show little evidence of racial disparity by their criteria of CPW, it was very interesting to see how the rest of their paper progressed.

The authors calculate Moran's I on model residuals, however given that model residuals are based off of an OLS model, it would be useful for the formal equation of the OLS model to also be provided in their methodology. What I personally really appreciated from this section was that given their Lagrange Multiplier Tests for both spatial error and spatial lag were statistically significant, they proceed to provide their own qualitative reasoning as to why they proceeded the rest of their analysis with a spatial error model. This demonstrates a good effort on the authors part to not just rely on empirical tests, but also to consider the context of spatial autocorrelation in police precincts. I especially liked their example of neighboring precincts all live near each other; hence their patrolling patterns may be similar. This was demonstrated as the "friends" example given by the instructor for this class. One thing to note however is that the same example may be applied for a spatial lag model as well. The authors note that there is equal justification for using a spatial lag model, however this statement might be better fitted later on in their discussion section. Lastly, it is through their formal equation of their spatial error model that the reader is introduced to the independent variables for the first time. It might serve useful to the reader if the independent variables were explicitly stated earlier in the introduction or methodology section. Currently, the reader doesn't know how the independent variables were selected or how it controls for confounding errors.

Small details of improvement for section 2.1 (Model Diagnostics)

- There is a formatting error when stating the assumptions for a spatial error model. I believe that the list of assumptions was meant to be either bullet pointed, or there is a comma between each assumption.
- The tense for checking assumptions suddenly shifts to past tense, while the rest of the paper is in present tense. Instead of saying "we checked" or "we examined" or "we calculated", it might be more consistent to instead say "upon checking" or "upon examining… we find" etc.

The authors state in section 2.2 that they also conducted a sensitivity analysis with a spatial lag model, however the entire paragraph is too short and brief at four lines. It states that "we compare the regression coefficients with those from the spatial lag model as well as those from the OLS model", but it is not indicated where these comparisons can be found (there is no reference to the Appendix). Additionally, the authors might benefit from moving the sensitivity analysis writeup to their discussion section, since it refers to statistically significant regression coefficients, but the reader has not gotten to the results section yet. Thus, the reader currently does not know which specific predictors were significant and how they are interpreted for the spatial error model, much less for the other two models. The authors might be able to add more substance to this section by stating the implications of the three models being similar in significance.

The Discussion section 4.1 'Interpretations' was succinct, and all coefficient interpretations were correct. Section 4.2 revisits again their sensitivity analysis, where they state that both spatial lag and spatial error are better than an OLS model given the significance of their autoregressive coefficient for both models. Again, it might be useful to the reader to see the results table for the spatial lag and OLS model, however the authors do not refer to the appendix. The authors may also benefit from expanding on their section 4.3 'Critique on Methodology and Alternative Model Specifications'. They add a citation stating that "majority White precincts are more likely to be low crime areas", but it would be interesting to know why such the case is. The authors could possibly mention majority White precincts being 'richer' neighborhoods, where both employment and education level are higher. Overall, if they added literature review for this section to make statements, it would be useful to be thorough and more comprehensive. Section 4.4 talks about 'Limitations', but to defend their paper it would also be great for the authors to talk about the strengths as well. Their confidence in their paper may appear weak especially if they start this section with the statement "This study's focus on weapon discovery is unusual". I could argue that it could have been a strength and not a weakness if they were to have strong literature review and justification for their focus beginning from the introduction of their report. Overall, team 08's paper was unique in its approach, but ultimately it might benefit from tying CPW more to the original aim of the study, which was to look for racial disparity in the severity of SQF events in NY. I agree with the authors' statement that "the results of this study may not be generalizable to SQF events overall". As mentioned before, weapon discovery might justify a police's intentions to stop a suspect, however given the context of American laws related to owning a weapon, weapon discovery is weakly tied to racial disparity. It may be that ethnic minorities live in more dangerous neighborhoods were safety is not as robust, thus they carry weapons as a means of self-protection (especially for women). Additionally, there is not enough statements on how discovery of weapons might have spatial correlations at a precinct level. I think a more explicit statements related to how neighborhood precincts might have similar socioeconomic levels, which is related to safety, thus police have higher incentive to stop suspects in these precincts might have served useful. The code for this repository was 100% reproducible and I ran into no errors.

[redacted]

STA 440 Case 3 Peer Review Team 8

The report "Case 3" by Mishek Thapa, Tong Wu, Sibora Seranaj, and Jessie Ou sets out to analyze whether there is evidence of racial disparities in Stop-Question-Frisk events in New York City after taking into account the spatial effects of precinct. Focusing on incidents in which the suspected crime was the criminal possession of weapons, the authors examine the precinct-level weapon recovery rate and relate it to the proportion of Black and Hispanic suspects stopped in a precinct and the proportion of the precinct's residents that are Black and Hispanic in order to examine racial disparities. Overall, their paper is thorough and convincing, although there are some opportunities to improve some details and wording as well as opportunities to clarify and bolster their argument.

In the introduction section, the authors provide background on the policy and previous research regarding SQF events. They clearly state their goal in the second paragraph of the paper, which is a strong choice because it allows the reader to understand the central argument of their paper from the start rather than having to read several pages before doing so. They provide a strong justification for focusing only on suspected weapon possession, since it allows them to circumvent many of the problems that previous studies using hit rates were criticized for. This simple but smart approach is a laudable strategy to examine racial disparities in one specific outcome. The authors also later (appropriately) acknowledge that this choice narrows the scope of the claims they can make from their data and could miss other types of racial disparities in SQF outcomes (as almost any study would). In another strong part of their introduction, they clearly underscore that their analysis can only study disparate outcomes and not discrimination, which is important to make clear in a statistical study since it is quite easy to interpret the results incorrectly. Descriptions of variables are concise but clearly worded, which is appreciated since it makes the paper easy to read and understand. The figures in their exploratory data analysis are professionally formatted and well labeled (although both titles could follow standard or at least the same capitalization conventions), and the captions are especially informative so that it is easy to understand what takeaways the reader should have from the inclusion of these figures. The introduction is an especially strong portion of this paper.

In the methodology section, their model is clear but there are a few minor errors and ways the authors could improve the flow of the section. First, the authors begin by talking about the residuals from a regression, but never actually specify what regression they are running. The reader can assume that they run the same regression as the one found later in the section but without the spatial dependence term. However, without it, this paragraph is abrupt and confusing, so adding a description of what this first model is would make the paper flow more readily (also, Moran's I is spelled incorrectly). A similar issue arises when the discuss Lagrange Multiplier tests on lambda and rho but do not explain what lambda and rho are or what models they are part of. The authors should include details about the exact tests they are performing so

that their methods are reproducible. They could also consider moving these hypothesis tests to the Appendix since they are not essential in building the central argument of the paper. In the model equation, one minor formatting change could be to include an indicator symbol to show that majority white is a binary variable. Although the authors provide reasons to include a spatial error term, what justifications do they have for not also adding a spatial lag term? Since authors show that there is statistical evidence supporting its inclusion, it would be helpful to explain why they do not do so, a question that naturally arises to the reader. They do include a spatial lag model as a robustness check, but the reasoning behind the decision to use spatial error as the main model and not to include a model with both terms could be further explained. There are some typesetting issues in Section 2.1 when discussing the assumptions for the spatial error model, which can be easily fixed in a future draft. The writing of the methodology section could use some work, but the methodology itself is well-justified and appropriate for answering the research question.

The results section provides a table of the estimates of the model and their uncertainty. The table is professionally formatted and easy to read. The authors could consider making a forest plot of their model estimates and confidence intervals, since this provides a clear and quickly digestible depiction of the magnitude of and uncertainty around the model estimates. The results section is simple and effective.

In the discussion section, the authors interpret their results, critique their methodology, and propose avenues for future work. The authors clearly state that they have failed to find evidence of a racial disparity of policing but also do not conclude that no such disparity exists, which is a principled, correct, and clear interpretation of their results. They propose using alternative geographic scopes of spatial data, which makes sense in the context of the "ecological fallacy," which warns that patterns among aggregate groups may not reflect the patterns at more granular levels and is frequently referenced when studying spatial inequalities. Finally, they note the limits of examining only weapons crimes and suggest that future studies could look at a broader picture of outcome disparities, which is an inherent trade-off when using this approach to avoid some of the problems associated with hit rates. Overall, the discussion section is logical, clear, and sensible.

The appendix is professionally formatted and the captions on the figure in Section 6.1 is highly informative, as was the case earlier in the paper. The authors could include additional text explain their plots in Section 6.2. It is not currently clear what takeaways they want from the reader. It would be helpful if in Section 6.3, the authors displayed the estimates from all three models side-by-side in one table. It is difficult to scroll back and forth between the results section and the appendix to compare the model outputs.

The code provided was quickly reproducible in under 10 minutes. No issues emerged when cloning their GitHub repository and compiling their RMarkdown document.

Overall, the paper provides a thorough analysis of one specific SQF outcome. The central decision of the paper, to focus on events based on suspected possession of its weapon, causes both its major strengths and weaknesses. First, it allows the authors to circumvent several of the issues surrounding hit rates. However, it significantly narrows the scope of the paper and uses only a small fraction of the information the authors have access to. There are some distracting errors or flow inconsistencies in the paper, especially in the methodology section, but these can easily be fixed in a later draft of the paper.

[redacted]

Case Team Reviewed: Team 8

Overall, the manuscript was decently written, and main sections of the paper were appropriate and effective. However, there were many minor issues and areas for improvement with the paper. I will go through each of the sections of the manuscript as they are presented, providing my thoughts on each section of the paper and make suggestions for improvement, if applicable. Then, I will discuss the overall presentation of the paper and strengths and limitations of the authors' analysis. Lastly, I will take the new, interesting, and relevant parts of this paper to reflect on and possibility to improve my own paper.

The introduction in the case study is somewhat effective. The authors provide relevant background information, with the appropriate citations, that tells a story and also explains why the study was conducted. Research questions and goals that the authors seek to evaluate are also clearly defined. I especially liked the high-level contextualization of the problem being studied. Although the data used in the study was well described, but descriptions of data manipulation was not included. Some of the variables included in the methodology section were not present in the original dataset, and some form of explanations either in the main body of the paper or in the Appendix would be appreciated and help the reader better understand the processes conducted by the authors.

Although the introduction and data section were well-written overall, some improvements could be made. Section 1.1 discusses the reasoning for studying CPW events, but parts of this could be brought up to the introduction to give better context prior to discussing the case study goals. In general, this section requires proof-reading, as tense errors and incorrect word forms were noticeable throughout this section (e.g., suspicious). In spite of these shortcomings, overall, this introduction serves as a decent gateway to the rest of the paper.

Figure 1 for the EDA somewhat fulfills its role in indicating where spatial clusters may occur. However, the caption was not very useful; perhaps specifying exactly where the clusters exist on the map would be beneficial. In addition, it seems that the outlier precinct is heavily skewing the gradient of the map; perhaps removing this outlier would help convey the differences between the precincts better.

Figure 2 of the EDA was nice in comparing weapon recovery between minority and white precincts. However, discussion of precincts 22 and 13 was confusing, as these precincts do not seem to have special significance via the histograms.

Overall, the methodology section presented appropriate reasoning and background for the spatial error model. I liked the discussion on potential confounders that were unaccounted for. The model is clearly specified with the appropriate mathematical notation and variable notations were defined clearly.

However, improvements could be made here for better clarity. Presentation of the multiple subsections in the methodology could be reordered for better readability. The methodology discusses model selection, then reasoning, and back to model selection. This is

confusing to read as a reader. In addition, more discussion of why the authors believed the residuals are correlated due to confounders would be beneficial.

The authors did not discuss the reasoning behind why specific predictor variables were selected at all. The authors could have referenced EDA figures and provided more background to help support the selection of certain predictor variables. In addition, some of the variables selected appear to be very similarly related, such as proportion of black residents to majority white. The author should have perhaps combined these variables into an interaction term.

The model diagnostics and various model assumptions are appropriate, along with the methods used to check assumptions, but results of the assumptions were not discussed. However, benefits of the spatial error model compared to other models such as the spatial lag model in answering the research question would have made the methodology more informative.

The authors mentioned justification for the spatial lag model, but did not discuss those in detail, nor why in spite of such justifications, they selected the spatial error model.

The results for the spatial error model are neatly presented in a clean and informative table, and all the variables stated from the model are presented.

The discussion does a good job in summarizing the results and the lack of significant predictors. However, providing a written interpretation for proportion of CPW suspects that were Black/Hispanic is strange, given that this variable was not significant. Limitations of the choice of response variable are mentioned but discussing limitations of the spatial error model itself on the lack of significant results could be helpful. The sensitivity analyses with other models were appropriate and reasonable.

Overall, the paper was well organized. Throughout most of the report, the paper employs a professional and formal tone. However, the paper cited resources through in-text citations, but the References were presented as links. This makes it difficult to determine exactly which specific source the text is referencing. If the References are just links, then it will be better if the citations just provided a number instead of the author names. The authors could benefit from reviewing the paper for grammatical errors or phrasing that could be improved upon to make the paper flow and read better. Specifically, it appears that the authors were trying to list the model assumptions in Section 2.1 on multiple lines but failed to do so. In addition, the authors mention ‘Moron I’ test in the methodology section. While amusing, such a test does not exist, and the authors likely meant ‘Moran’s I’. These mistakes could have been caught by a thorough read over of the report. However, the discussion, presentation, and explanation of concepts is presented in a manner that is easily understandable for peers with similar statistical backgrounds.

The approach that the authors took is similar to part of the methodology in which my team took in this analysis, so nothing was new or interesting to me. The choice to make a variable of majority white or not as a binary variable was an interesting choice. Although I would not necessarily implement this change immediately, I would definitely plot a figure to see the relation between this and arrest rate (our response variable), and then conduct an ANOVA to test the significance of this variable on our model.

The analysis was very easily reproduceable. From downloading the GitHub repository as a ZIP file, then unzipping, opening it as a new R project, finding and opening the final report, to knitting to pdf, took less than 2 minutes. All the plots and figures used in the final report were included in the final report markdown itself, so no additional downloads were required.

[redacted]

Case Team Reviewed: Team 8

This case study presents a decent analysis of how, at the precinct level, weapon discovery rates in SQF events where police suspect the criminal possession of a weapon is not significantly associated with the proportion of Black or Hispanic suspects in such events. However, there several clear areas of improvement that the authors could pursue. While the introduction offers a succinct background and motivation for the research goals, the research goals are confusing. The phrase "taking into consideration the spatial effects of precinct" implies that precinct imparts some spatial effect on the response variable. Also, the authors mention a model for the "weapon recovery rate", only to discuss how they expect the "weapon discovery rate" would be lower in precincts with higher proportions of Black and Hispanic suspects. The authors could dispel this confusion by discussing how they want to account for "spatial effects across precincts", and change "weapon recovery rate" to "weapon discovery rate". The authors also delve into a discussion of their approach, a section that would be more appropriate for the methodology section.

The authors provide detailed explanations of their data sources, and they define their response variable clearly, adding EDA that clearly shows the spatial dependencies in the response variable they want to account for. This EDA is effective in justifying why the authors suspect spatial dependencies in the weapon discovery rate by precinct. However, any discussion of what predictor variables will be used is missing. Because of this, it is unclear exactly what the authors are planning to compare the weapon discovery rate to and what confounding variables the authors are planning to control for.

In the methodology, the authors outline their approach clearly. However, because the p-values both Lagrange multiplier tests were statistically significant, it would be helpful to see if more robust tests produce the same results. The model formulation is clear, but because of the lack of discussion about predictor variables, it is somewhat unclear what the variables refer to. For example, the majority white variable presumably refers to whether or not the precinct has a majority of white people, but it could also refer to whether or not the suspects in SQF events where criminal possession of a weapon is suspected for the precinct are majority white. The caption for Figure 2 in the Exploratory Data Analysis section clears up this confusion, but it would be helpful to put such a distinction in the main body of the text. The authors do discuss what they plan to do to check their model assumptions and diagnostics, but when they list out the assumptions they want to check, they forget to add commas, which makes the list confusing. They also fail to summarize their findings in the main body, which would have saved the reader time from having to check the appendix to see if the model assumptions and diagnostics turned out to be reasonable. The authors have a similar problem with their sensitivity analysis, where they discuss the use of a spatial lag model with the same covariates, but they don't mention any key findings.

The results authors obtain is put into a table that is formatted well. One small improvement the authors could make is to put parentheses around each confidence interval since it is standard to have them. The authors also make a point about "when weapon discovery rate in neighboring precincts increases, so does the weapon discovery rate in each precinct". This isn't worded too well since it seems to imply that increases in the weapon discovery rate in neighboring precincts will make the weapon discovery rate in every precinct rise. If the authors state "For each precinct, when weapon discovery rate in neighboring precincts increases, so does

the weapon discovery rate in that precinct", the statement would make more sense and communicate the finding more effectively.

In the discussion section, the authors offer a faulty interpretation of the results. They state that "for every 0.01 increase…the expected weapon discovery rate decreases by 0.049", which is mathematically incorrect. It should be "for every 1 increase…the expected weapon discovery rate decreases by 0.049". Aside from this issue, the discussion of the results was sensible. The discussion of limitations and of future avenues of work are also reasonable, and the authors raise good points about extending the analysis to other types of crime.

Overall, the case study is well presented and professional. The figures and tables are clear and effective. One thing the authors could afford to do is to check their grammar so that the narrative of the report is more clear. The appendix section is well-organized, and the flow of tables and figures is sensible. Reproduction of the case report mostly went smoothly. There was an issue where a package in the Rmd would not install on a Duke VM, but luckily the report does not depend on the package. The whole process of reproducing the report took about 10 minutes.