**1.** I think what the creator is intending to convey is the regions with the highest densities of homicides. It looks like they used kernel density interpolation that extends about 2000 ft out in all directions, so the highest density value points (lightest colors on this map) are the weighted distance sums of all other points with nearer points contributing at a greater magnitude (and points farther than 2000 ft having minimal or no effect).   As a result, the point on the map with the highest value/lightest color does not contain any incidents of homicide but is surrounded by multiple small, but dense clusters whose combined effects out total any individual cluster.

2. The original patterns my group saw, influenced by the fact we could not see super clearly from a distance, was that homicides appeared to be clustered alongside the major roads depicted in white. We were stumped about why and thought maybe it had something to do with how crimes are reported, or criminals' attempts at hiding bodies and evasion of arrest.   We also thought the roads could correspond to gang territories and so these homicides are reflective of gang conflict and violence particularly regarding territory.

3. I apologize, but this topic really caught my interest, and I ended up going down the rabbit hole. So, for the ease of grading I'm to split my answer into two sections. The first one just being summary thoughts and the second all my research, questions, and feelings.

a. I noticed several inconsistencies and potential problems with the graph after looking at the source website and the FBI's UCR site.

For one, the threshold scale for hotspots is being defined based on the local range of values and not a standardized, comparable scale. Because homicides are rare and the dataset small, it gives a false impression of the intensity of hot spots. "Clusters" of only two or three are registering when realistically small clusters are a natural result of random distribution. The author did not include any results of statistical significance testing so it's impossible for us to conclude these spatial relations are significant or just noise. Example Graphs Source
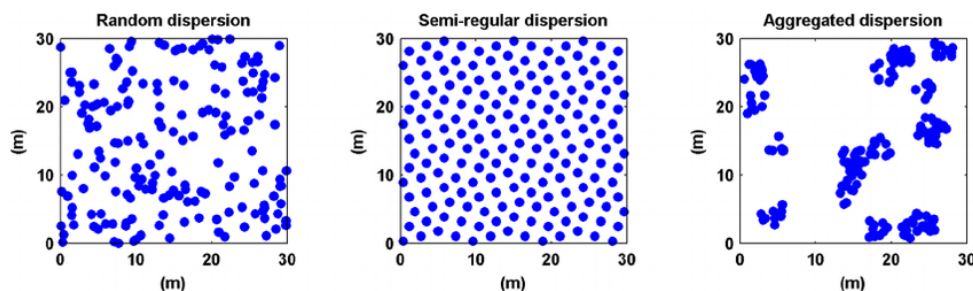
I think the author should have defined a threshold prior to mapping in determining what is a meaningful cluster density. What that exact value should be, I don't know, but looking at general trends of past data and considering the intent or application of the produced map. For example, if the intent is for

preventative hotspot policing, then setting the effect radius of each incident to 300 feet might be preferred because it better aligns with an individual's visual capacity which is a key component of prevention by police presence. Or saying a hot spot region must contain a minimum of 5 homicides.

For this map, I think the underlying processes are a combination of the above-mentioned questionable analysis and more general problems spatial criminology such as fuzziness of limiting locations to the spaces of defined addresses or categorizing crimes by type as definitions can vary across time and agencies and many incidents are comprised of several types (example: theft by assault with an illegally possessed firearm could fit in three categories).

b.

Critique of Data Visualization & Communication

I first looked at the original source for the graph and was surprised at the quality of data science and communication.   The map presented has a sister map of homicides in 2021 (note also that it is only January to April)

ABQ Homicides
January 1, 2021 - April 11, 2021
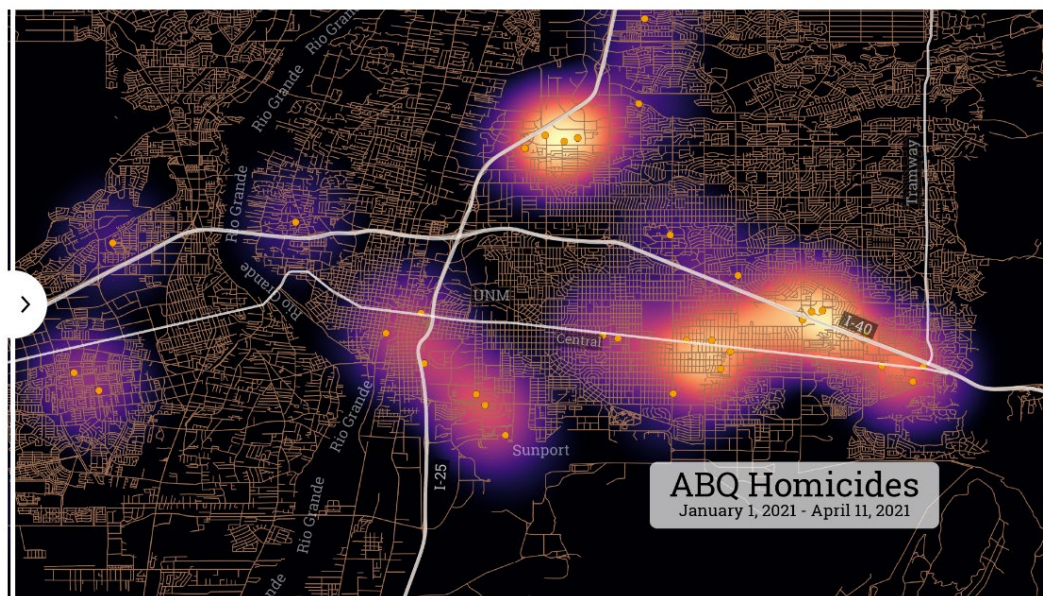
Noticeably, the hot spot classifications are not on the same scale, yet they are presented within the same interactive slider and the caption notes "homicides have been happening farther east." All of this prompts the audience to make direct comparisons between the maps, but since the hot spot threshold is relative to each year's total, points in the smaller 2021 data set individually have a larger magnitude of effect in defining hotspots. In the 2020 map, the beginnings of a hotspot (dark purple) do not begin to show until clusters of two or three. Comparatively, the 2021 map places a visible hot spot gradient around every single incident even those whose closest neighbors are more than half a mile away.

Additionally, when I pulled up a map of ABQ to estimate the distances mentioned in Question 1, I noticed Albuquerque's city boundaries are quite irregular, shown in the following left side image). So, I found a map of APD's district and reporting area (pictured below on right colored by APD area command regions.) which proves that entire sections of populated city development are not included in the crime data yet are included in the density calculations. The implications of this heterogeneity of space are that it skews the probability equations for significance by underestimating the expected crime at a given point or area. Another curious area is the small, excluded block in the orange section 1 right where the other maps had placed the hottest spot.
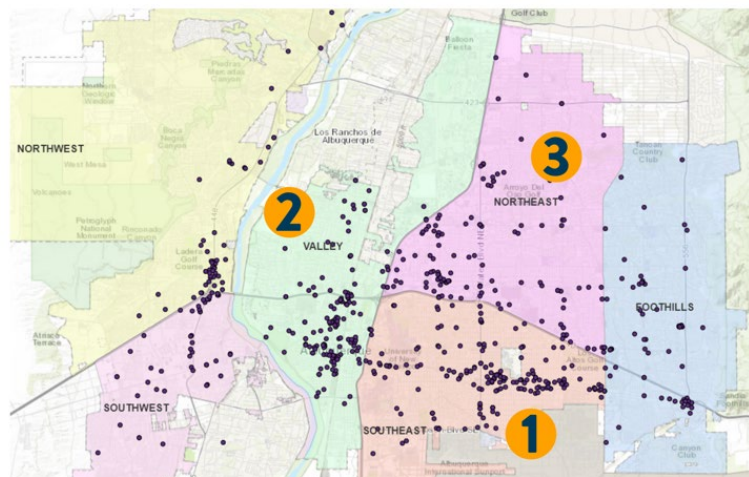
History & Conceptual Foundations of Crime Hot Spots

I also spent some time working through some of the hallmark literature on crime hot spots and policing. Modern theories and applications of it are based on two twentieth century studies:

1. Kansas City Preventive Patrol Experiment: A Technical Report | Office of Justice Programs (ojp.gov)

This 1974 Kansas City experiment concluded that doubling police presence in neighborhoods does not reduce or prevent crime.

2. The Minneapolis Hot Spot Patrol Experiment (1995)

This experiment, which took place in the late 1980s and published in 1995 specifically set out to disprove the Kansas City Experiment. Just from my own observations, this study is used as the foundation for almost every subsequent hot spot research or evaluation, just look for the in-text citation (Lawrence and Weisburd, 1995). It found that police presence did decrease the relative number of public calls to dispatch and visible disordered conduct.

Honestly, this study is worth reading for anyone in the social sciences because of the sheer number of questionable practices and conclusions in experimental design, implementation, and statistical analysis of which I cannot possibly cover all here, but if anyone else reads this please share your thoughts with me. I would love to really explore some of these details with someone and see if we have the same conclusions.

Questionable Statistics?

I do want to comment on the fact that the statistics were specifically done to give them the largest possible chance of finding a statistically significant result. They did this by applying a one-tailed mixed methods ANOVA test with alpha=0.1 or 10%.

If you are not familiar with these terms, it is similar to (and note I'm extremely oversimplifying here) a statistical significance level of $p < 0.2$ for the more common two-tailed test.   For reference, many sciences require $p < 0.05$ to reject the null hypothesis and some fields like clinical medicine require $p < 0.01$.

There is nothing inherently wrong with this as it minimizes type II errors (false negatives), but at the cost of risking type I errors (false positives). Balancing the risk of each type of error depends on the research topic and intended applications. For example, fraud detection algorithms prefer false positives over false negatives which would mean failing to identify and stop fraudulent activity in time. On the other hand, variable relationships which have a small magnitude of effect will be almost impossible to detect at a defensible level of significance for publication and results supporting the null hypothesis rarely get published.

Having read the report, I do wonder if the choice of minimum $p < 0.10$ was not made after the calculations as the majority of findings would not meet the standard $p < 0.05$ criteria even with the one-tailed advantage. Of course, this is only speculation and ultimately p-values are only a measure of probability against a hypothetical sampling distribution and not a magic number below which all findings are true and above which all findings are false. But the researchers did at one point conflate significance with magnitude of effect so I have no idea what these researchers actually did to get these values.

Additionally, it was unclear what categories were used for the mixed methods ANOVA. I would assume it was a typical repeated samples analysis where the intragroup variable is each month of the study, but they also talk a lot about these hierarchical "blocks" or groupings of hot spots that are paired with a similar block in the control group so its possible the intragroup categories were these blocks. In which case, they've significantly reduced their number of observations in an already small study. I would question the amount of expected error in the result and