

1 Formatting Tips

When you are writing a link in RMarkdown you can suppress the full URL by writing '[text to show](https://someFullURL.com)'. Your report pdf renders each link twice but links are rendered correctly in the html. In the interactive component section too, there appear to be four links for the one web app.

When you are writing an RMarkdown list if you end each line with two spaces, then it will be rendered as a line break. For example, in your conclusion the bulleted list of variables rendered as '-Suspect's Sex -Reason for ...', instead of as a list like you intended.

2 Explanations

Your arrests by neighborhood map is impressive! In your description, there are a few times when you reference the 'density' of arrests, but I am unsure that is the correct term when writing about count data. If you are not normalising these counts by neighborhood populations, ie frequency / population, then you might want to refer to the metrics as counts or frequencies rather than densities.

I am not sure what is gained by plotting the random sample of arrest locations in leaflet. Your explanation was that it serve as a tool for predicting future crime locations. If that is the intention then maybe there could be an explanation about crimes frequently reoccurring in the same location.

When you are referencing neighborhoods, as on page 13 'Rochdale, Springfield Gardens, ...', it might be hard for a non-New Yorker to figure out which areas of the graph to look for. Perhaps you could look into including pointers to specific neighborhoods that you would like to reference. If this is difficult in ggplot then editing your images in a dedicated editing program could simplify the process.

3 Organization

I would suggest that you leave all explanation of future direction and limitation until the conclusion. Discussing that you would like to have performed additional analyses is a bit distracting. However, I completely understand the urge to qualify anything you are even the slightest bit unsure of. You might try to ask yourself what it is that you are sure of, eg 'Blacks are arrested more than whites', what you suspect, eg 'Blacks are arrested disproportionately more than whites', and try to reconcile how you can prove the latter. If you can't then I think it is fine to stick to what you know and later explain that the limitations and how you want to go about addressing them.

Including a feature about the proportion of the population for each race would give context to the arrest counts. I appreciate the reference to demographics in your Presentation section, but perhaps you could include demographics in your feature engineering. For example, finding the relative frequency of arrests

by race by share of the NYC population. In your 2010 NYC census race results, double encoding race in color, eg black mapped to purple, white-hispanic mapped to blue, ... might make it easier to recognize that the ordering of race on the x-axis changes from that of the stop-and-frisk chart.

I would have liked to see some of the exploratory data analysis work that went into coming up with the graphs in your Presentation section, eg the data on efficacy and on felony classification.

In your interactive Stop-and-Frisk rates map, with the sliding scale for years, perhaps you could include a color scale for the map. I understand that it is a similar diverging color scale as in the previous choropleth zip code maps but including a scale here would allow the web app to function as a stand-alone tool. If it is too hard to include in D3, then a static image of the viridis scale with frequencies could serve as a replacement.

In your conclusion section, I appreciated the explanation of limitations and future areas of interest. Perhaps you could have included some brief summary of your important findings from the data exploration and presentation here, right before the limitations.

4 Ideological Note

When writing about the prevalence of blacks among complaints and shootings it is worth noting that the data you have comes from the NYPD, an organization with its own agenda and limitations. This is in contrast to the true rates of crime, which may or may not be accurately represented by data on arrests and complaints. This is a fact about data more generally so I understand if you did not see fit to include referency this explicitly, but since you are talking about race and policing I thought it was worth a second thought. Kudos to you for taking on such an important topic at all! Not only was the data messy but you must have known that people cannot help but bring their personal beliefs with them even when looking at data. I think you made a good effort to stick to the data while providing a bit of context when writing about the transition from the Bloomberg-era to De Blasio.

In future analyses, you could look for data from cities in which a stop-and-frisk policy was not implemented. The “broken windows” style of policing really came to prominence in the 80s, but NYC under Rudy Giuliani brought it to national prominence. Looking for data in other American cities could be hard. It might prove impossible if you wanted to find a city with similar demographics to NYC, ie in terms of size and diversity, but with a different policing strategy. European cities would not have adopted broken windows but also might have demographic differences that make it hard to draw parallels with NYC.