

Lab 3 Slides- Challenge 2 and 4

By Andrew White.

Challenge 2- Problem and Dataset

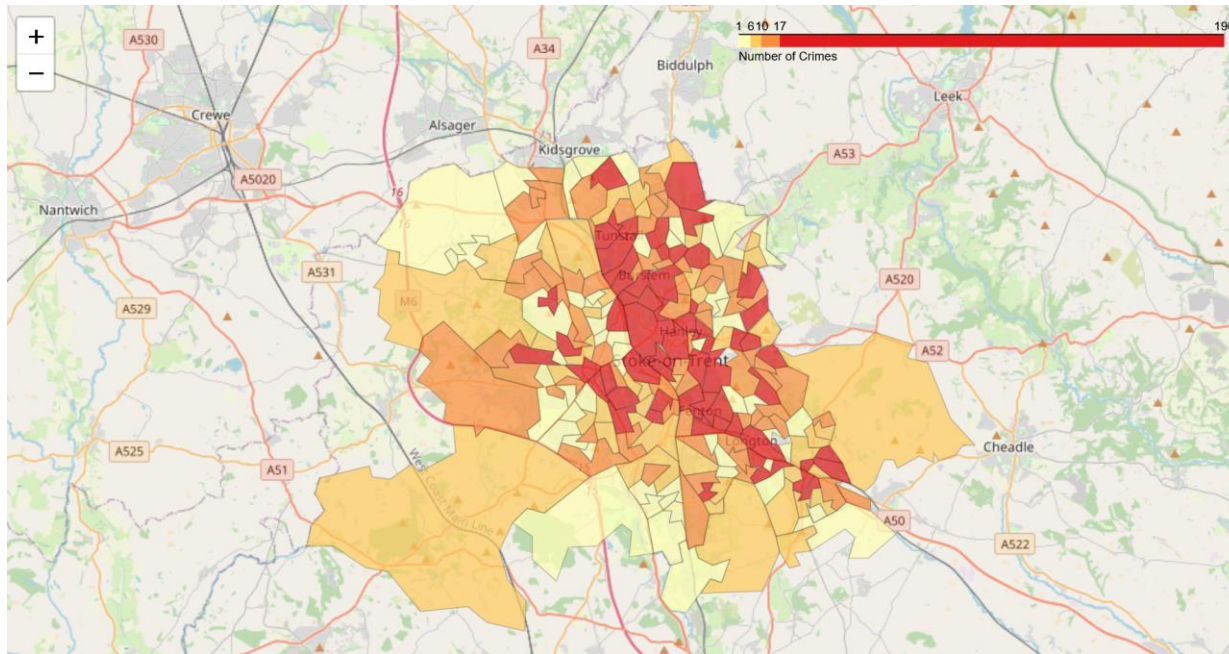
- The problem that I wanted to explore was street crime in Stoke on Trent.
- To do this, I used the Police crime API with a custom polygon with boundaries around Stoke on Trent.
- I decided to find the total number of instances of each committed crime type in Stoke on Trent (2024-12), the number of each crime type under investigation and the ratio of under investigation to total crimes of each type.
- Statistics for the ratio of under investigation to total crimes were gathered, number of investigations per crime type charts were generated and maps created (finding crimes per LSOA and folium heat maps).

Challenge 2- Outcomes and Challenges

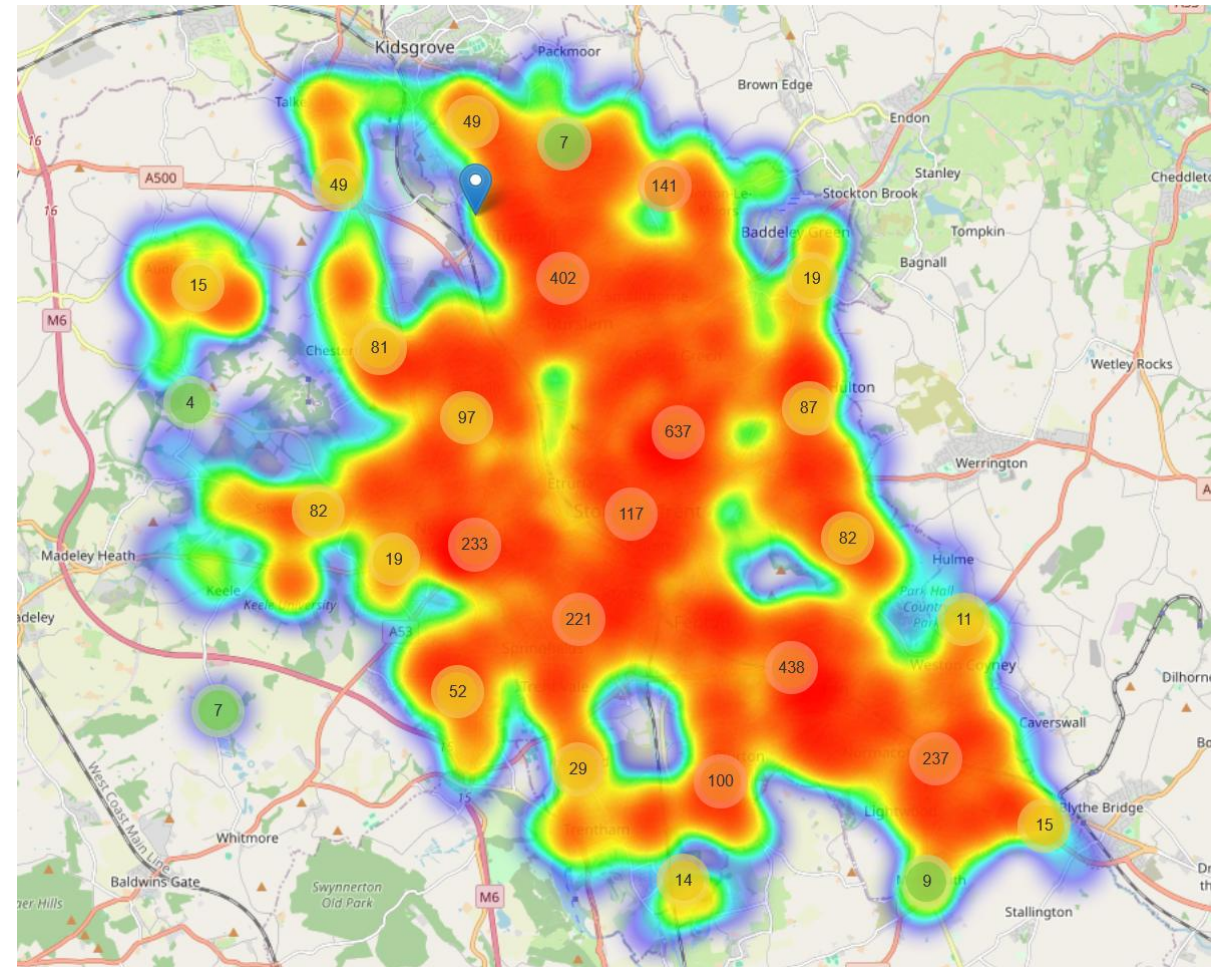
- From this, I found that different crimes had different ratios of investigation afterward.
- Whereas crimes like Anti Social Behaviour and bike theft has low ratios of investigation, more serious crimes like robbery and violent crimes were more likely to be under investigation.
- The mean ratio for under investigation across all crimes was 0.52, median was 0.55 and standard deviation was 0.18.
- One of the biggest challenges for this dataset was separating out the variables to extract latitude and longitude values from the location. I had to use `item.get()` in order to get the information, via `item.get('location', {}).get('latitude')` (first, finding the location value then, the latitude.)

Appendix 1- maps for challenge 2

Crimes per LSOA



Crime Heatmap



Challenge 4- Problem and Dataset

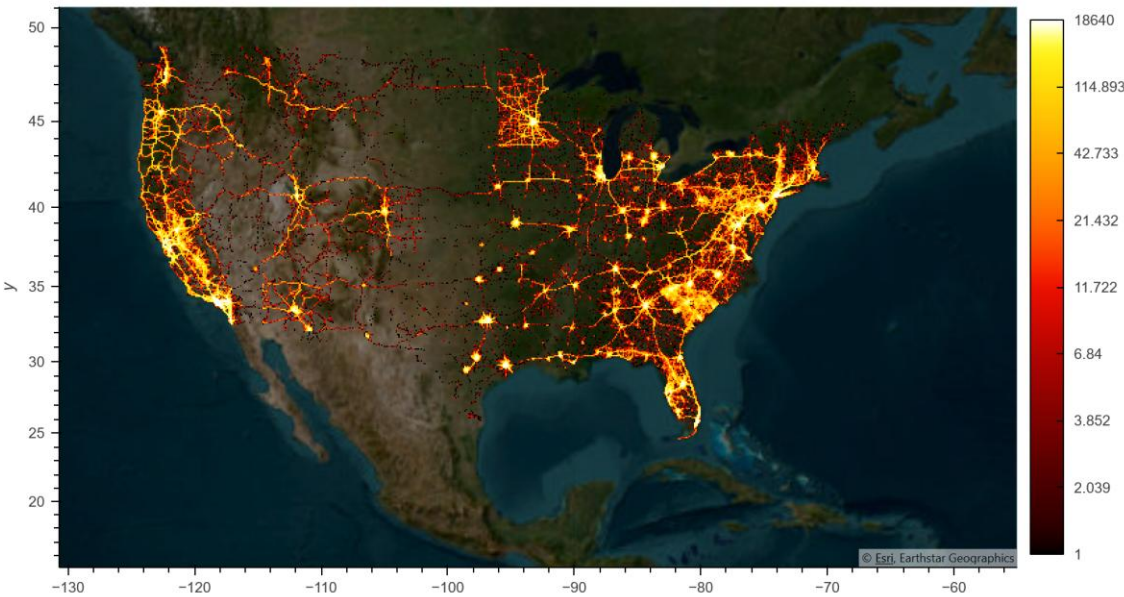
- For this map, the problem that I was identifying was car accidents in the continental United States. I got the dataset from Kaggle (<https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents?resource=download>). This dataset contained 7.7 million variables between Feb 2016 and March 2023.
- I wanted to see the number of incidents specifically at night so, I filtered out all the accidents during the daytime.
- I also wanted to plot the severity of accidents. To do that I used the severity information given (1 is least severe, 4 is most severe).
- The dataset needed to be cleaned up (all NA values removed) before use.

Challenge 4- Outcomes and Challenges

- From the data, I found that on average, the severity of incidents was higher in the Eastern United States compared to the Midwest and Western states.
- However, as the dataset combined multiple different datasets together, some states recorded severity differently. Notably, you can make out outlier states such as South Carolina which mark most incidents as severity one.
- A challenge that I had with plotting the data was using the correct co-ordinates and CRS. By default, the data was gathered as GPS co-ordinates however, Datashader required the use of web Mercator co-ordinates (in meters). This meant that I needed to use the `set.crs` and `to.crs` functions to reproject the co-ordinates so they could be plotted.

Appendix 2- map for challenge 4

Map of road accidents at night between
Feb 2016 and March 2023



Severity of road accidents at night
between Feb 2016 and March 2023

