**Intro**

In my final project I will be assessing whether the number of burglaries within key city’s is correlated to the distance of the property to what is considered a primary road, for example a motorway junction. The idea being investigated is that burglars are more likely to target properties and areas that have excellent highway connections to a primary road allowing them to both get away quickly as well as disguise themselves amongst a large amount of traffic making it less likely the police will be able to identify their vehicle in a sufficient time. The output areas I have chosen for this investigation are prominent cities in the North of England; Leeds, Bradford, Sheffield, and Hull.

The Census data was taken from a bulk download of 2021 from the NOMIS webpage (Nomis, 2021). The chosen category’s which would have the most influence was TS006: Population density, TS017: Household size and TS050 number of bedrooms. As each category is its own excel sheet, I elected to merge all of them together before preprocessing using python. I also formatted the household size into a simple two columns; properties with up to three people and properties more than 3 people, reducing from the original eight columns.

The Crime data was downloaded using a custom data download widget, specifying February 2023 – February 2024, this was the latest data available and I wanted to analyse the data over a one-year period (data.police, no date). This download was separated by month so I combined all twelve sheets into one spreadsheet. Originally, I was going to convert the excel file into a geo-frame within python, however I was not able to read the geo-frame in further command lines so I opted to add the excel file to QGIS and create the geometry, saving as a shapefile. Since I was creating a new shapefile, I also converted the results into British National Grid (BNG) from latitude longitude.

I downloaded the road network data from Digimaps using the BNG tile function, SE, SK, and TA (digimap, 2023). As each of these contained 2 separate shapefiles for network and nodes, I merged all the Road\_Link shapefiles together however this was a large file at 500MB so I decided to condense this further before processing in python by selecting all roads that intersect inside or out, the target cities. In addition to the road data, I also created an ad-hoc standalone shapefile containing the starting points of all the roads where primary equals true i.e. Motorways, major trunk roads or arterial roads.

Output areas (polygons) were downloaded from Geoportal (Geoportal Statistics, 2023), I went with the Dec 2021 download to match as close as possible the recorded census data as the LSOA codes generally has slight variation between years. I filtered this output using the website interactive mapping tool and drew a box around my area of interest. This allowed me to download a much smaller area than the original entire country’s worth of areas, I also manually edited this in QGIS to further remove extraneous polygons that were not in the study areas making the large file size as small as possible.

**Preprocessing:**

I clean the Output areas to only include the chosen city’s using a string contains function to find key words and text as the output areas download does not have a standalone column with city names by default. Using the explore function you can visualise the area of my investigation clearly, this file with be used as an important template for other datasets later. Checking the coordinate systems are all the same.

I create a dissolved polygon of the city’s and then clipped the road links shapefile using this polygon to only use the relevant roads located in and adjoining the cities. I chose to include the adjoining roads as well as there may be instances where roads are on the boundary of the city but not in which would alter the estimated road distance quite drastically in some cases. Creating a dissolved polygon works better than 6,000 individual ones as some roads may be clipped in unexpected ways, this also helps with the speed and optimisation of the python file.

My original plan was to upload the crimes excel sheet and convert into a geodata frame using the Fiona package however I experienced problems with the folium map section (Data Ninjas, 2021). So, I opted instead to create the shapefile as above.

I was also going to calculate the distance between points along a network to gain a true distance from burglary to primary road but due to a problem of conflicting tool instructions and package version instructions I could not get the get\_nearest\_node command to work (automating-gis-processes, 2017). To get around this I chose to use the Euclidian distance (do-me, 2023), although this would not be a true account it could still be used as a reasonable estimate of distance for this use case.

Utilising the spatial join command against burglary points and the primary road start points I created my new data frame, using a maximum distance tolerance of 2000m to reduce any outlier distances as most of the city would be covered within 2000m. I also calculated some additional statistics using this distance information to join onto the LSOA data frame using the group-by and aggregate functions to summarise the count of burglaries and the average distance to the primary road per output area.

For the last part of data processing, I split the data frames into their respective city limits Leeds, Sheffield, Bradford. Hull was left out of the final visualisation as there were no burglary points visualised in the Hull area.

**Visuals**

Using Spearman’s correlation, I found that the average distance has an overall weak relationship to the number of burglaries but a moderate relationship between number of burglary and both one- and two-bedroom accommodation. These are found on the outskirts of the city, however there are also some accommodations within. For this visual I used a mask to cover off the other half of the data to tidy up the map. I used the RdBu colour palette as using colour brewer as a reference of different possible palettes for my graphs and this one is a colourblind friendly choice and is best used as a diverging colour scheme (Aptech, 2024).

For my visuals I chose a green colour for the joint plot scatter graph, this was going to be blue but I though the contrast between the dark blue and light blue colours made it difficult to clearly see the results so I opted for a green instead, a single colour tone will not affect visually impaired users. Using the bar charts, you can see a cluster near the 250m mark which steadily decreases with greater distances from the primary roads.

Better visualised by the hex plot graph it is far easier to spot clusters of significance in darker blue where there is lots of overlap of results, which was very difficult to identify on the previous graph. The correlation shows a larger concentration of burglary’s along the 250m mark which carries over to 500m and steadily declines from the 750m mark. I also renamed the axis labels to be more intuitive for the users who will be the Police force and crime analysts giving them a clear and easily read graph.

For the final visual I plotted the spatial representation of burglary and added the primary road layer relative to the output areas allowing for easy identification of hotspots and clusters of data. This is so the police and crime analysts will instinctively be able to tell areas of focus against the road network their familiar with to reduce the crime rate. I used the ‘YlOrRd’ colour palette as this is both colour blind friendly (Aptech, 2024) and you can easily distinguish the different level of burglary rates by using the ‘temperature’ of the colour.

On the maps, you can see a high concentration near the centre of Leeds which is completely encircled by primary roads and other clusters directly in between these prime roads. Lighter colours are seen farther from the prime roads. This description can be effectively applied to all the other maps as well.

**References**

Aptech. 2024. *Color Brewer Palettes.* [Online]. [Accessed 28 April 2024]. Available from: https://www.aptech.com/releases/gauss18/graphics-updates/color-brewer-palettes/

Tableau. 2024. *5 tips on designing colour-blind friendly visualisation.* [Online]. [Accessed 26 April 2024]. Available from: https://www.tableau.com/en-gb/blog/examining-data-viz-rules-dont-use-red-green-together

Data Ninjas. 2021. *Create a shapefile form a .csv with Python, web map it and reproject it.* [Online]. [Accessed 20 April 2024]. Available from: https://www.youtube.com/watch?v=RAn8Lx0qrks

jeremysze. 2019. *Network Distance.* [Online]. [Accessed 29 April 2024]. Available from: https://jeremysze.github.io/GIS\_exploration/build/html/networkdistance\_osm.html

automating-gis-processes. 2017. *Network analysis in Python.* [Online]. [Accessed 29 April 2024]. Available from: https://automating-gis-processes.github.io/2017/lessons/L7/network-analysis.html

Aric A. Hagberg, Daniel A. Schult and Pieter J. Swart, [*“Exploring network structure, dynamics, and function using NetworkX”*](https://conference.scipy.org/proceedings/SciPy2008/paper_2/), in [Proceedings of the 7th Python in Science Conference (SciPy2008)](https://conference.scipy.org/proceedings/SciPy2008/index.html), Gäel Varoquaux, Travis Vaught, and Jarrod Millman (Eds), (Pasadena, CA USA), pp. 11–15, Aug 2008

do-me. 2023. As of v0.10.0 geopandas supports sjoin\_nearest natively. 13 October 2023.[*Finding nearest point in other GeoDataFrame using GeoPandas*](https://gis.stackexchange.com/questions/222315/finding-nearest-point-in-other-geodataframe-using-geopandas)*.* [Online]. [Accessed 26 April 2024]. Available from: https://gis.stackexchange.com/questions/222315/finding-nearest-point-in-other-geodataframe-using-geopandas

nomis. 2021. TS006 Population Density. *Office for National Statistics*. [Online]. [Accessed 7 April 2024]. Available from: https://www.nomisweb.co.uk/sources/census\_2021\_bulk

nomis. 2021. TS0017 Household Size. *Office for National Statistics*. [Online]. [Accessed 7 April 2024]. Available from: https://www.nomisweb.co.uk/sources/census\_2021\_bulk

nomis. 2021. TS050 Number of Bedrooms. *Office for National Statistics*. [Online]. [Accessed 7 April 2024]. Available from: https://www.nomisweb.co.uk/sources/census\_2021\_bulk

Geoportal Statistics. 2023. Output Areas (December 2021) Boundaries EW BGC V2*. ONS Geography*. [Online]. [Accessed 7 April 2024]. Available from: https://geoportal.statistics.gov.uk/datasets/6beafcfd9b9c4c9993a06b6b199d7e6d\_0/explore?location=53.594232%2C-1.209805%2C9.00

data.police. no date. Data Download. *data.police*. [Online]. [Accessed 9 April 2024]. Available from: https://data.police.uk/data/

digimap. 2023. Highways – Roads (GB). *Digimap Edina*. [Online]. [Accessed 9 April 2024]. Available from: https://digimap.edina.ac.uk/roam/map/os