Crimes in the UK

An awesome Data Mining Project

[**1. Phase 1: Crimes Data only**](#_h0cxs3grlprj)[**3**](#_h0cxs3grlprj)

[1.1. Original Data Sets](#_1oichm9sqdnf) [3](#_1oichm9sqdnf)

[1.1.1. data.police.uk](#_tfe1pdbndglp) [3](#_tfe1pdbndglp)

[Cleanings](#_vddx64r2vo97) [3](#_vddx64r2vo97)

[Data Attributes and Statistics](#_3ema32q3nrgq) [4](#_3ema32q3nrgq)

[Streets](#_lm0vb78cyrhh) [4](#_lm0vb78cyrhh)

[Outcomes](#_g2qatpweirb6) [5](#_g2qatpweirb6)

[Stop and Search](#_sbxakfxmu695) [5](#_sbxakfxmu695)

[1.1.2. Points of Interest](#_j4rvomfg4i9q) [6](#_j4rvomfg4i9q)

[1.1.3. Geographical Data](#_qnzx27vwceka) [7](#_qnzx27vwceka)

[1.2. Combining Data Sets](#_f2k2y4gfhhss) [8](#_f2k2y4gfhhss)

[1.2.1. Crime-centric](#_wtdu3ghvyuwl) [8](#_wtdu3ghvyuwl)

[1.2.2. Location-centric](#_js47xqo78zjz) [10](#_js47xqo78zjz)

[1.3. Data Availability](#_cerjbbmdtx7f) [12](#_cerjbbmdtx7f)

[1.3.1. Countries](#_ed4z1v4b2hw) [12](#_ed4z1v4b2hw)

[1.3.2. Police Forces](#_qlu1w3h594gy) [13](#_qlu1w3h594gy)

[1.3.3. Darkmap](#_1p5d39qsoaqk) [14](#_1p5d39qsoaqk)

[1.4. Stop and Searches](#_udvig8f435mr) [15](#_udvig8f435mr)

[1.4.1. Simple Statistics](#_axo3wtknj769) [15](#_axo3wtknj769)

[Age Range](#_7hfb98axr73g) [15](#_7hfb98axr73g)

[Ethnicity](#_kdoqhnog4zol) [15](#_kdoqhnog4zol)

[Gender](#_dvzwhryjcgoi) [16](#_dvzwhryjcgoi)

[Over the Time](#_pk7okilgkch1) [16](#_pk7okilgkch1)

[1.5. Crimes](#_tcuhd6cr3t9u) [17](#_tcuhd6cr3t9u)

[1.5.1. Simple Statistics](#_j1ldgxs642) [17](#_j1ldgxs642)

[1.5.2. Crime Types](#_mz8uwhrus9u8) [18](#_mz8uwhrus9u8)

[1.5.3. Outcomes](#_1nuvde4mmh3p) [20](#_1nuvde4mmh3p)

[1.6. Geographical Distributions](#_t108fe1nk9uq) [26](#_t108fe1nk9uq)

[1.6.1. Total Number of Crimes in the Regions](#_74ol0sx72wua) [26](#_74ol0sx72wua)

[1.6.2. Total Number of Crimes in the Postcode Areas](#_kzbfrvhktjm0) [27](#_kzbfrvhktjm0)

[1.7. Random Forest](#_m7xp1kpihgrb) [27](#_m7xp1kpihgrb)

[**2. Phase 2: Demographical Influences**](#_fvwt3eq0qwtk)[**28**](#_fvwt3eq0qwtk)

# 

# 

# 1. Phase 1: Crimes Data only

In the first phase of our project, we focused on the crimes data. …

## 1.1. Original Data Sets

### 1.1.1. data.police.uk

Our main data set is provided by data.police.uk. It is structured in 70 folders. Each folder represents the crime data for a single month of the respective year, broken down by territorial police force. The source files are stored as .CSV and have 13.2 GB. There are three types of data: Street-level crimes (called “*streets*”), street-level outcomes (called “*outcomes*”) and *stop and search* information.

The Street-level crimes data ranges from december 2010 till September 2016, the street-level outcomes data ranges from January 2012 till September 2016, and the stop and search data ranges from December 2014 till September 2016.

For privacy reasons, the data is being anonymized. For this, the location points of the data have been mapped to more than 750.000 map points all around UK. The generated data points represent some Points Of Interest (POI). If the nearest point is more than 20km away, the coordinates of the crimes are being removed. Also other fields like id or date have been anonymized.

#### Cleanings

In outcomes, there are 2547177 “crime\_id”s that exist more than one time. In streets, there are 145124 such cases. The tuples with the same “crime\_id” are very similar, so that we assume that they concern the same crime and that they accrued because of updates. Therefore, we removed the older tuples. In total, we removed 656261 tuples in streets and 3405907 tuples in outcomes.

We also removed all tuples before September 2011 because in September 2011, a lot of new types of crimes were introduced. We deleted 5069641 tuples in streets.

We dropped the column “reported\_by” in outcomes, because it has always the same value like “falls\_within”. Moreover, we dropped “context”, because the values are not very meaningful and is null in 99.64% of the cases.

Streets and outcomes have beside of the coordinates-columns also a column named “location”, that contains a small description of a place. In some cases, the same place have different location values, because some values are different spelled (e.g. “On or near Moretons Lane” and “On or near Moreton'S Lane”). We corrected the differences.

In the data sets from September 2013, “lsoa\_code” is sometimes missing although a corresponding lsoa does exist and is given in other tuples with the same coordinates. We corrected it.

#### Data Attributes and Statistics

After cleanings, we have the following attributes and number of tuples.

##### Streets

The streets relation has 29289398 tuples. The table below describes each attribute in the “streets” files of the dataset.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Attribute name** | **Missing values** | **Number of distinct values** | **Example value** | **Comment** |
| crime\_id | 11758313 (40.1%) | 17531085 | 42a1d4194f0a7d0f7f6e0d4e709c7f2d68a321e2b28f9a8fc5203c329bbdc5a4 |  |
| year\_month | 0 (0%) | 61 | 2015-09 |  |
| reported\_by | 0 (0%) | 45 | Hertfordshire Constabulary | Except 82506 cases, it is the same like falls\_within. |
| falls\_within | 0 (0%) | 45 | Hertfordshire Constabulary |  |
| longitude | 355661 (1.2%) | 749319 | -0.57560600 |  |
| latitude | 355661 (1.2%) | 720917 | 51.76402600 |  |
| location | 355661 (1.2%) | 275005 | On or near BECKETTS |  |
| lsoa\_code | 699241 (2.4%) | 34750 | E01023370 | LSOA are only given for crimes in England and Wales. Scotland and Ireland have similar geographical areas[[1]](#footnote-2), but they are not given by this data set. |
| lsoa\_name | 699241 (2.4%) | 34750 | Dacorum 012B |  |
| crime\_type | 0 (0%) | 16 | Anti-social behaviour |  |
| last\_outcome \_category | 12823779 (43.8%) | 25 | Unable to prosecute suspect |  |
| police\_force | 0 (0%) | 45 | hertfordshire | A variant spelling of falls\_within, that is used for joining with stop and search. |

##### Outcomes

The outcomes relation has 14218059 tuples. The table below describes each attribute in the “outcomes” files of the dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Attribute name** | **Missing values** | **Number of distinct values** | **Example value** | **Comment** |
| crime\_id | 0 (0%) | 14218059 | 42a1d4194f0a7d0f7f6e0d4e709c7f2d68a321e2b28f9a8fc5203c329bbdc5a4 |  |
| year\_month | 0 (0%) | 57 | 2015-09 |  |
| police\_force | 0 (0%) | 43 | hertfordshire | See above |
| falls\_within | 0 (0%) | 43 | Hertfordshire Constabulary |  |
| longitude | 257468 (1.8%) | 648721 | -0.57560600 |  |
| latitude | 257468 (1.8%) | 630170 | 51.76402600 |  |
| location | 257468 (1.8%) | 253550 | On or near BECKETTS |  |
| lsoa\_code | 257924 (1.8%) | 34749 | E01023370 | See above |
| lsoa\_name | 257924 (1.8%) | 34749 | Dacorum 012B |  |
| outcome\_type | 0 (0%) | 23 | Offender given a caution |  |

##### Stop and Search

The stop and search relation has 584841 tuples. The table below describes each attribute in the stop and search files of the dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Attribute name** | **Missing values** | **Number of distinct values** | **Example value** | **Comment** |
| type | 0 (0%) | 3 | Person search |  |
| year\_month\_day | 0 (0%) | 278644 | 2015-09-01 |  |
| part\_of\_a\_policing \_operation | 410668 (70.2%) | 2 | False |  |
| policing\_operation | 569290 (97.3%) | 1316 | OPERATION BLACKJACK |  |
| latitude | 274057 (46.8%) | 97563 | 53.73099100 |  |
| longitude | 274057 (46.8%) | 98148 | -1.75077100 |  |
| gender | 57273 (9.7%) | 3 | Male |  |
| age\_range | 33856 (5.7%) | 5 | 18-24 |  |
| self\_defined \_ethnicity | 8979 (1.5%) | 17 | White - White British (W1) |  |
| officer\_defined \_ethnicity | 48310 (8.2%) | 5 | White |  |
| legislation | 47621 (8.1%) | 14 | Police and Criminal Evidence Act 1984 (section 1) |  |
| object\_of\_search | 113984 (19.4%) | 14 | Article for use in theft |  |
| outcome | 17957 (3%) | 14 | Nothing found - no further action |  |
| outcome\_linked\_to \_object\_of\_search | 445630 (76.1%) | 2 | True |  |
| removal\_of\_more \_than \_just\_outer\_clothing | 426419 (72.9%) | 2 | False |  |
| police\_force | 0 (0%) | 44 | west yorkshire |  |

### 1.1.2. Points of Interest

As an enrichment to the crimes data, we used a data set[[2]](#footnote-3) that provides around 191,000 POI all over the UK, broken down by the different categories listed below.

* Wifi hotspot
* Fuel: Gas station
* Shopping consumer electronics: electrical retailer selling home electronics and household appliances
* Transport: airports and train stations
* Accommodation: Hotels
* Entertainment: Casinos, Cinemas
* Banks
* Food and drink: restaurants, dining pups
* Government agency: Tax office, passport office
* Public service buildings: Police station, fire station
* Landmark: Mountains, hills, waterfalls
* Attraction: Aquarium, museums
* Book shop
* Special Interest: Madame Tussaud, Church bell
* Community: churches, Universities
* Courier: DHL, Express
* Healthcare: Hospital, practice
* Sports center: sports clubs
* Pharmacy
* Shopping supermarkets: Lidl, Aldi
* Association
* Holiday parks
* ATM

### 1.1.3. Geographical Data

In addition to the coordinates, we also had the Lower Layer Super Output Areas (LSOA) code of the area where a crime happened. LSOAs have an average population size of around 1,500 citizens. This low granularity was i) too small to be visualized on a map in an informative manner ii) not enough to discover location-based patterns. Therefore, we integrated more geographical data[[3]](#footnote-4). This allowed us to come up with the following geographical structure and therefore derive new features about the crime locations:

|  |  |
| --- | --- |
| **Area Type** | **Description** |
| Latitude/Longitude | The anonymised coordinates. (~800.000 values) |
| Lower Layer Super Output Area (LSOA) | Small areas defined by the Office for National Statistics with about 1,500 citizens. (~35,000 values) |
| Middle Layer Super Output Areas (MSOA) | Like LSOA but with about 7,200 citizens. (~7,200 values) |
| Postcode Area | There are 121 postcode areas in the UK. We have data about crimes in 105 areas. |
| Region | England can be divided into 9 regions. Furthermore, we also consider Wales, Scotland and Northern Ireland as regions. |

## 1.2. Combining Data Sets

### 1.2.1. Crime-centric

Using the three data sets, we created two relations. The first one is crime-centric and called *crimes*.

In order to detect the nearest stop and search incidents and POIs within the range of 500m from crimes, we used PostGIS, a spatial database extender for PostgreSQL. Because of performance reasons, we used as distance metric the L-infinity norm, so we could build indices with the coordinates of Stop-and-Search and Points-of-Interest points. Afterwards, we were able to perform a fast search of surrounding points for every crime entry - which still took many hours on our local machines.

The *crimes* relation has the following attributes:

* id
* crime\_id
* year\_month
* reported\_by
* falls\_within
* latitude
* longitude
* location
* lsoa\_code
* lsoa\_name
* msoa\_code
* msoa\_name
* local\_authority\_code
* local\_authority\_name
* postcode\_area
* region\_code
* region\_name
* **crime\_type**
* last\_outcome\_category
* suspect\_identified
* police\_force
* outcome\_type
* suspect\_identified\_outcome
* number\_500m\_ss
* number\_500m\_ss\_person\_searches
* number\_500m\_ss\_vehicle\_searches
* number\_500m\_ss\_asian\_searches
* number\_500m\_ss\_black\_searches
* number\_500m\_ss\_white\_searches
* number\_500m\_ss\_outcome\_suspect\_arrested
* number\_500m\_ss\_outcome\_article\_found
* number\_500m\_ss\_outcome\_offender\_given\_drugs\_p\_warning
* number\_500m\_ss\_outcome\_offender\_cautioned
* number\_500m\_ss\_outcome\_suspect\_summonsed\_to\_court
* number\_500m\_ss\_outcome\_nothing\_found
* number\_500m\_ss\_outcome\_offender\_given\_penalty\_notice
* number\_500m\_ss\_outcome\_local\_resolution
* number\_500m\_ss\_age\_under\_10
* number\_500m\_ss\_age\_10\_17
* number\_500m\_ss\_age\_18\_24
* number\_500m\_ss\_age\_25\_34
* number\_500m\_ss\_age\_over\_34
* number\_500m\_ss\_object\_controlled\_drugs
* number\_500m\_ss\_object\_criminal\_demage\_articles
* number\_500m\_ss\_object\_stolen\_goods
* number\_500m\_ss\_object\_weapons
* number\_500m\_ss\_object\_firearms
* number\_500m\_ss\_object\_evidence
* number\_500m\_poi
* number\_500m\_poi\_community
* number\_500m\_poi\_food\_and\_drink
* number\_500m\_poi\_fuel
* number\_500m\_poi\_atm
* number\_500m\_poi\_transport
* number\_500m\_poi\_motorcycle\_dealership
* number\_500m\_poi\_wifi\_hotspot
* number\_500m\_poi\_shopping\_supermarkets
* number\_500m\_poi\_special\_interest
* number\_500m\_poi\_shopping\_others
* number\_500m\_poi\_banks
* number\_500m\_poi\_sports\_center
* number\_500m\_poi\_car\_dealership
* number\_500m\_poi\_commercial\_site
* number\_500m\_poi\_association
* number\_500m\_poi\_car\_parts\_supplier
* number\_500m\_poi\_builder\_and\_diy\_supplies
* number\_500m\_poi\_accommodation
* number\_500m\_poi\_attraction
* number\_500m\_poi\_pharmacy
* number\_500m\_poi\_healthcare
* number\_500m\_poi\_landmark
* number\_500m\_poi\_public\_service\_buildings
* number\_500m\_poi\_shopping\_consumer\_electrics
* number\_500m\_poi\_entertainment
* number\_500m\_poi\_car\_servicing
* number\_500m\_poi\_building\_society
* number\_500m\_poi\_bookshop
* number\_500m\_poi\_truck\_dealership
* number\_500m\_poi\_holiday\_park
* number\_500m\_poi\_government\_agency
* number\_500m\_poi\_vehicle\_hire
* number\_500m\_poi\_car\_auction
* number\_500m\_poi\_caravan\_dealership
* number\_500m\_poi\_courier
* number\_500m\_poi\_estate\_agent

### 1.2.2. Location-centric

Our second approach is location-centric and called locations. The relation *locations* has latitude/longitude entries as rows. Therefore one row represents crimes data about one location. The table could be used to analyze crime trends and anomalies in different type of locations.

The locations table has the following attributes:

* latitude
* longitude
* location
* lsoa\_code
* lsoa\_name
* msoa\_code
* msoa\_name
* local\_authority\_code
* local\_authority\_name
* postcode\_area
* region\_code
* region\_name
* number\_500m\_ss
* number\_500m\_ss\_person\_searches
* number\_500m\_ss\_vehicle\_searches
* number\_500m\_ss\_asian\_searches
* number\_500m\_ss\_black\_searches
* number\_500m\_ss\_white\_searches
* number\_500m\_ss\_outcome\_suspect\_arrested
* number\_500m\_ss\_outcome\_article\_found
* number\_500m\_ss\_outcome\_offender\_given\_drugs\_p\_warning
* number\_500m\_ss\_outcome\_offender\_cautioned
* number\_500m\_ss\_outcome\_suspect\_summonsed\_to\_court
* number\_500m\_ss\_outcome\_nothing\_found
* number\_500m\_ss\_outcome\_offender\_given\_penalty\_notice
* number\_500m\_ss\_outcome\_local\_resolution
* number\_500m\_ss\_age\_under\_10
* number\_500m\_ss\_age\_10\_17
* number\_500m\_ss\_age\_18\_24
* number\_500m\_ss\_age\_25\_34
* number\_500m\_ss\_age\_over\_34
* number\_500m\_ss\_object\_controlled\_drugs
* number\_500m\_ss\_object\_criminal\_demage\_articles
* number\_500m\_ss\_object\_stolen\_goods
* number\_500m\_ss\_object\_weapons
* number\_500m\_ss\_object\_firearms
* number\_500m\_ss\_object\_evidence
* number\_500m\_poi
* number\_500m\_poi\_community
* number\_500m\_poi\_food\_and\_drink
* number\_500m\_poi\_fuel
* number\_500m\_poi\_atm
* number\_500m\_poi\_transport
* number\_500m\_poi\_motorcycle\_dealership
* number\_500m\_poi\_wifi\_hotspot
* number\_500m\_poi\_shopping\_supermarkets
* number\_500m\_poi\_special\_interest
* number\_500m\_poi\_shopping\_others
* number\_500m\_poi\_banks
* number\_500m\_poi\_sports\_center
* number\_500m\_poi\_car\_dealership
* number\_500m\_poi\_commercial\_site
* number\_500m\_poi\_association
* number\_500m\_poi\_car\_parts\_supplier
* number\_500m\_poi\_builder\_and\_diy\_supplies
* number\_500m\_poi\_accommodation
* number\_500m\_poi\_attraction
* number\_500m\_poi\_pharmacy
* number\_500m\_poi\_healthcare
* number\_500m\_poi\_landmark
* number\_500m\_poi\_public\_service\_buildings
* number\_500m\_poi\_shopping\_consumer\_electrics
* number\_500m\_poi\_entertainment
* number\_500m\_poi\_car\_servicing
* number\_500m\_poi\_building\_society
* number\_500m\_poi\_bookshop
* number\_500m\_poi\_truck\_dealership
* number\_500m\_poi\_holiday\_park
* number\_500m\_poi\_government\_agency
* number\_500m\_poi\_vehicle\_hire
* number\_500m\_poi\_car\_auction
* number\_500m\_poi\_caravan\_dealership
* number\_500m\_poi\_courier
* number\_500m\_poi\_estate\_agent
* **number\_crimes**
* **number\_public\_order**
* **number\_anti\_social**
* **number\_bulgary**
* **number\_other\_crime**
* **number\_violence\_sexual\_offences**
* **number\_bicycle\_theft**
* **number\_drugs**
* **number\_criminal\_demage**
* **number\_other\_theft**
* **number\_shoplifting**
* **number\_vehicle\_crime**
* **number\_violent\_crime**
* **number\_public\_disorder\_weapon**
* **number\_robbery**
* **number\_theft\_from\_person**
* **number\_possessions\_weapons**

## 1.3. Data Availability

data.police.uk is a quite new platform, which composes data of several police forces all over the UK. These police forces entered the project at different times and use various systems so that the data from the different geographical areas are not available in the same volume.

### 1.3.1. Countries

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Country** | **Number location points** | **Number crimes** | **1st crime entry** | **Number stop and searches** | **1st stop and search entry** |
| England | 732558 | 27121189 | 2011-09 | 297031 | 2014-03-31 |
| Wales | 44829 | 1470012 | 2011-09 | 11295 | 2014-09-30 |
| Scotland | 326 | 11370 | 2011-09 | 3 | 2016-02-08 |
| Northern Ireland | 41315 | 331166 | 2011-09 | 90 | 2014-12-08 |

From Scotland, we only have data from British Transport Police, that is responsible for the railways in Great Britain.

### 1.3.2. Police Forces

### 

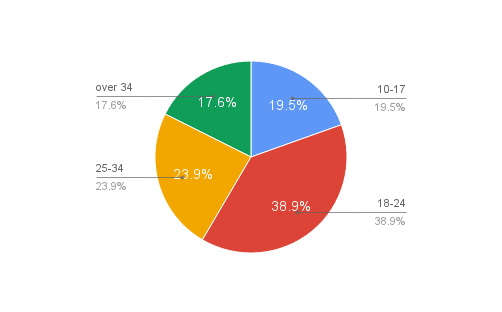
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Police force** | **Number location points** | **Number crimes** | **Number stop and searches** | **1st stop and search entry** |
| Avon and Somerset Constabulary | 23406 | 823502 | 6125 | 2014-12-01 |
| Bedfordshire Police | 7491 | 313498 | 3203 | 2015-03-31 |
| British Transport Police | 2781 | 236302 | 5048 | 2015-04-01 |
| Cambridgeshire Constabulary | 11827 | 375010 | 2645 | 2015-07-31 |
| Cheshire Constabulary | 16767 | 460259 | 7288 | 2015-03-31 |
| City of London Police | 602 | 33953 | 2018 | 2015-03-02 |
| Cleveland Police | 8566 | 437957 | 4716 | 2014-12-01 |
| Cumbria Constabulary | 8030 | 232653 | 3657 | 2015-07-31 |
| Derbyshire Constabulary | 16502 | 510179 | 3918 | 2015-04-01 |
| Devon & Cornwall Police | 25332 | 668174 | 8103 | 2015-03-01 |
| Dorset Police | 10686 | 340816 | 9630 | 2014-12-01 |
| Durham Constabulary | 8761 | 319035 | 5427 | 2015-03-31 |
| Dyfed-Powys Police | 7957 | 195835 | 6218 | 2015-01-01 |
| Essex Police | 22331 | 803825 | 4518 | 2015-04-01 |
| Gloucestershire Constabulary | 9253 | 280632 | 2544 | 2015-04-01 |
| Greater Manchester Police | 40968 | 1671091 | 9102 | 2015-01-01 |
| Gwent Police | 8932 | 305207 | 1933 | 2014-09-30 |
| Hampshire Constabulary | 26882 | 887366 | 33886 | 2014-03-31 |
| Hertfordshire Constabulary | 14146 | 451034 | 11817 | 2015-03-01 |
| Humberside Police | 12804 | 443441 | 2868 | 2014-12-01 |
| Kent Police | 22655 | 810707 | 11512 | 2015-03-31 |
| Lancashire Constabulary | 24010 | 898381 | 7535 | 2015-03-31 |
| Leicestershire Police | 14239 | 418823 | 3268 | 2014-12-01 |
| Lincolnshire Police | 12424 | 304192 | 5954 | 2015-01-01 |
| Merseyside Police | 19064 | 777752 | 14676 | 2015-05-31 |
| Metropolitan Police Service | 93131 | 5001436 | 224533 | 2015-03-31 |
| Norfolk Constabulary | 15245 | 352927 | 8064 | 2015-01-01 |
| North Wales Police | 11516 | 317687 | 2951 | 2014-12-01 |
| North Yorkshire Police | 13704 | 354853 | 7943 | 2014-12-01 |
| Northamptonshire Police | 10862 | 396922 | 2598 | 2015-02-01 |
| Northumbria Police | 18599 | 803885 | 9726 | 2015-03-01 |
| Nottinghamshire Police | 16145 | 558848 | 4615 | 2014-12-01 |
| Police Service of Northern Ireland | 41307 | 331427 | 0 | - |
| South Wales Police | 17762 | 663060 | 9919 | 2015-03-31 |
| South Yorkshire Police | 18997 | 937120 | 5338 | 2015-03-01 |
| Staffordshire Police | 16786 | 494211 | 6414 | 2015-01-01 |
| Suffolk Constabulary | 12109 | 320018 | 4333 | 2015-01-01 |
| Surrey Police | 16705 | 462186 | 10924 | 2014-12-01 |
| Sussex Police | 21708 | 734380 | 11829 | 2015-03-01 |
| Thames Valley Police | 30549 | 844882 | 16627 | 2015-03-31 |
| Warwickshire Police | 8197 | 245611 | 5184 | 2014-12-02 |
| West Mercia Police | 18856 | 579486 | 13652 | 2014-12-01 |
| West Midlands Police | 28219 | 1275494 | 31246 | 2014-12-01 |
| West Yorkshire Police | 35982 | 1330066 | 27540 | 2014-12-01 |
| Wiltshire Police | 10264 | 285275 | 3796 | 2015-01-01 |

We have data from all police forces since December 2010 except from Police Service of Northern Ireland. The first crime entries from them are from September 2011.

## 1.4. Stop and Searches

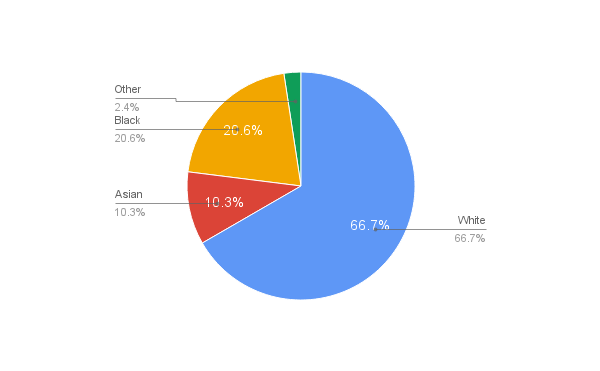
### 1.4.1. Simple Statistics

#### Age Range

The pie chart below describes the age range distribution in the stop and search table. Through this chart, it can be observed that there are far more people in the 18-24 age range that are being stopped and searched than in any other age range category. Except the 18-24 age range, the number of stop and searches by age range is similarly distributed.

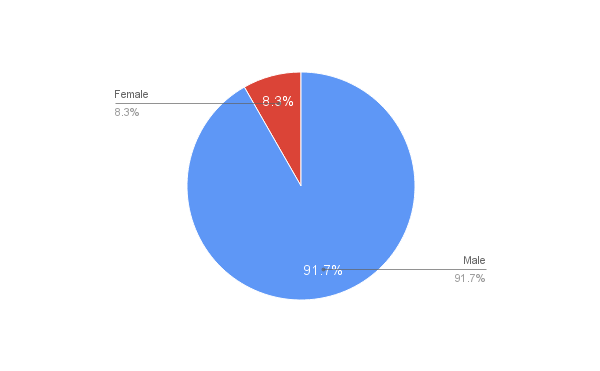
#### Ethnicity

The pie chart below describes the ethnicity distribution in the stop and search table. Through this chart, it can be observed that the most stopped and searched persons are white. However, these numbers do not express anything if not compared to the ethnicity distribution in the area where the stop and search has been made.



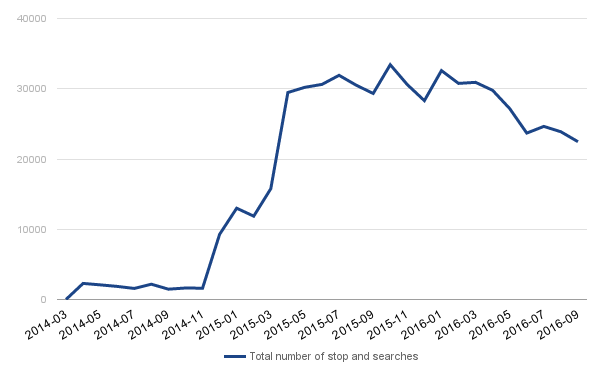
#### Gender

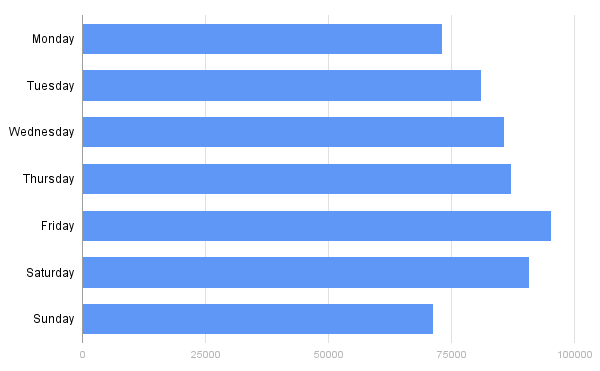
The pie chart below describes the gender distribution in the stop and search table. It is surprising to see that 9 out of 10 stopped and searched persons are males.



#### Over the Time

The time series graph below describes the amount of stop and searches that have been done until September 2016. It can be observed that the first year of this project is not very useful as data is missing. The second chart shows the days distribution on crimes. The crimes are similarly distributed among all week days.





## 1.5. Crimes

### 1.5.1. Simple Statistics

The time series below depicts the total number of crimes over time. In this graph, it can be observed that during winter there are typically less crimes than in the rest of the year. Conversely, the highest number of crimes is always in the summer

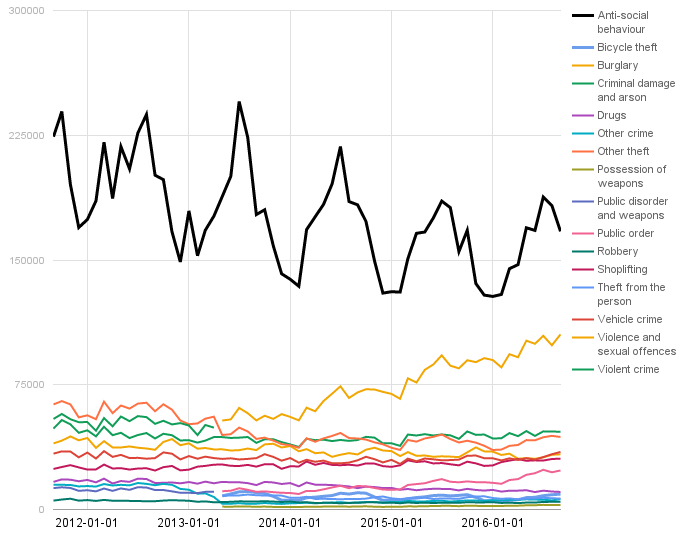
#### 

#### 

### 1.5.2. Crime Types

The crimes are divided in 16 categories[[4]](#footnote-5). This section analyzes the crime types and their evolution. The table below contains for each crime type the total number of entries, the date of the first entry, and the date of the last entry.

|  |  |  |  |
| --- | --- | --- | --- |
| **Crime type** | **Number crimes** | **1st entry** | **Last (given) entry** |
| Anti-social behaviour | 10723897 | 2011-09-01 | 2016-09-01 |
| Bicycle theft | 315680 | 2013-05-01 | 2016-09-01 |
| Burglary | 2184796 | 2011-09-01 | 2016-09-01 |
| Criminal damage and arson | 2674455 | 2011-09-01 | 2016-09-01 |
| Drugs | 881852 | 2011-09-01 | 2016-09-01 |
| Other crime | 462642 | 2011-09-01 | 2016-09-01 |
| Other theft | 2886095 | 2011-09-01 | 2016-09-01 |
| Possession of weapons | 80052 | 2013-05-01 | 2016-09-01 |
| Public disorder and weapons | 233368 | 2011-09-01 | 2013-04-01 |
| Public order | 603667 | 2013-05-01 | 2016-09-01 |
| Robbery | 281452 | 2011-09-01 | 2016-09-01 |
| Shoplifting | 1631288 | 2011-09-01 | 2016-09-01 |
| Theft from the person | 291188 | 2013-05-01 | 2016-09-01 |
| Vehicle crime | 1886783 | 2011-09-01 | 2016-09-01 |
| Violence and sexual offences | 3104835 | 2013-05-01 | 2016-09-01 |
| Violent crime | 1047348 | 2011-09-01 | 2013-04-01 |

[[5]](#footnote-6)

The time series above show the evolution of the number of crimes for each crime type. As it can be seen, the crime type “Public disorder and weapons” was splitted into “Possession of weapons” and “Public order” in April 2013. We want to have data for the whole period of time. Therefore we had to merge some of the columns. We decided to merge the splitted columns again to “Public disorder and weapons”. Furthermore, the crime type “Violent crime” was removed by the police.uk since May 2013. Instead, “Violence and sexual offences” was introduced. Again, we decided to merge them and name them as “Violence and sexual offences”.

After doing the cleanings described above, we have the following 13 crime types:

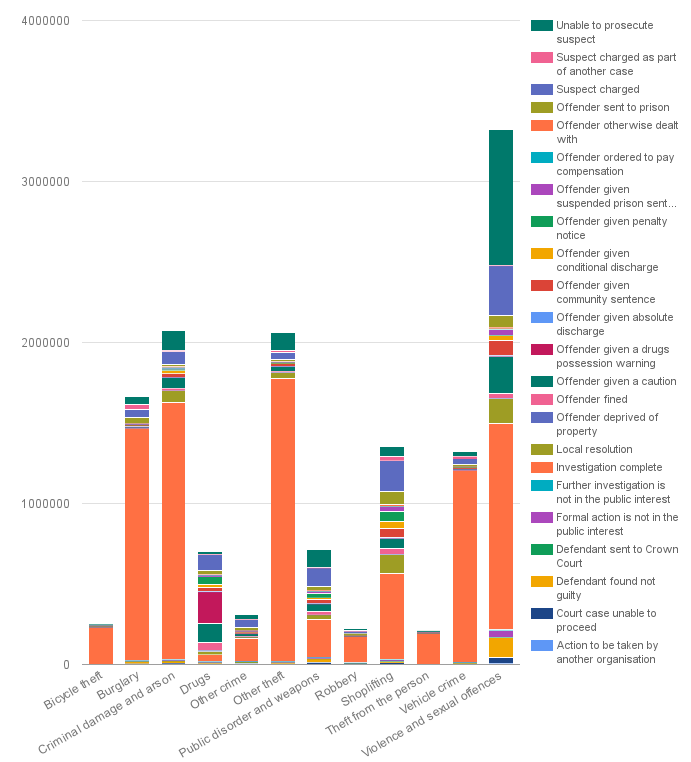
|  |  |  |  |
| --- | --- | --- | --- |
| **Crime type** | **Number crimes** | **1st entry** | **Last (given) entry** |
| Anti-social behaviour | 10723897 | 2011-09-01 | 2016-09-01 |
| Bicycle theft | 315680 | 2013-05-01 | 2016-09-01 |
| Burglary | 2184796 | 2011-09-01 | 2016-09-01 |
| Criminal damage and arson | 2674455 | 2011-09-01 | 2016-09-01 |
| Drugs | 881852 | 2011-09-01 | 2016-09-01 |
| Other crime | 462642 | 2011-09-01 | 2016-09-01 |
| Other theft | 2886095 | 2011-09-01 | 2016-09-01 |
| Public disorder and weapons | 917087 | 2011-09-01 | 2016-09-01 |
| Robbery | 281452 | 2011-09-01 | 2016-09-01 |
| Shoplifting | 1631288 | 2011-09-01 | 2016-09-01 |
| Theft from the person | 291188 | 2013-05-01 | 2016-09-01 |
| Vehicle crime | 1886783 | 2011-09-01 | 2016-09-01 |
| Violence and sexual offences | 4152183 | 2011-09-01 | 2016-09-01 |

### 1.5.3. Outcomes

Anti-social behavior does not appear in the outcomes table because of some difficulties that are described on the police.uk website[[6]](#footnote-7). There are 24 outcome types[[7]](#footnote-8).

|  |  |
| --- | --- |
| **Outcome type** | **Number crimes** |
| Action to be taken by another organisation | 10283 |
| Court case unable to proceed | 98287 |
| Defendant found not guilty | 221406 |
| Defendant sent to Crown Court | 1597 |
| Formal action is not in the public interest | 93106 |
| Further investigation is not in the public interest | 2328 |
| Investigation complete; no suspect identified | 8815109 |
| Local resolution | 453389 |
| Offender deprived of property | 6480 |
| Offender fined | 162119 |
| Offender given a caution | 585943 |
| Offender given a drugs possession warning | 208388 |
| Offender given absolute discharge | 5283 |
| Offender given community sentence | 276617 |
| Offender given conditional discharge | 130177 |
| Offender given penalty notice | 155306 |
| Offender given suspended prison sentence | 115366 |
| Offender ordered to pay compensation | 18202 |
| Offender otherwise dealt with | 38270 |
| Offender sent to prison | 305281 |
| Suspect charged | 1004765 |
| Suspect charged as part of another case | 91758 |
| Unable to prosecute suspect | 1414560 |

The next chart depicts the amount of outcome types per crime type.

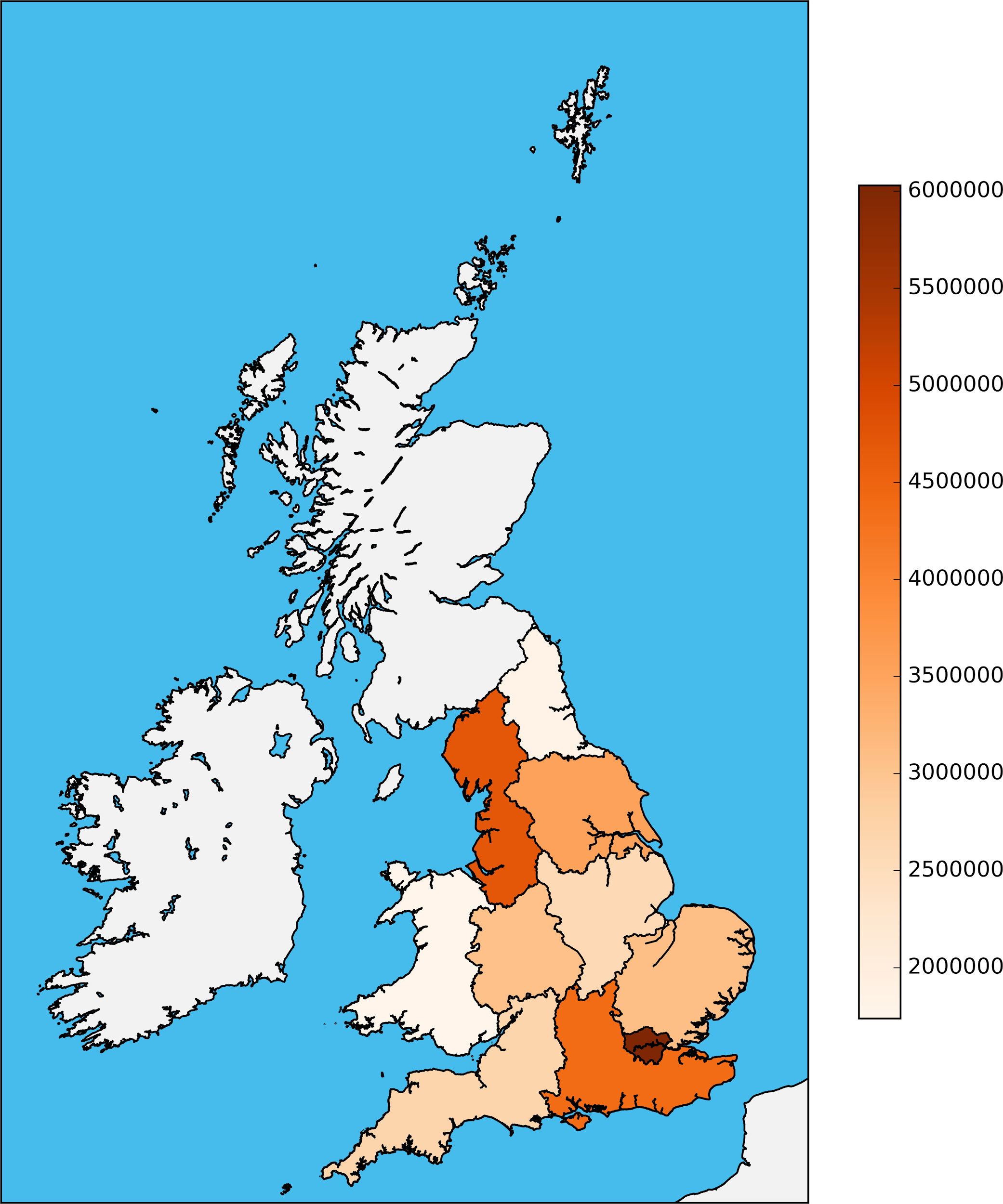


## 1.6. Geographical Distributions

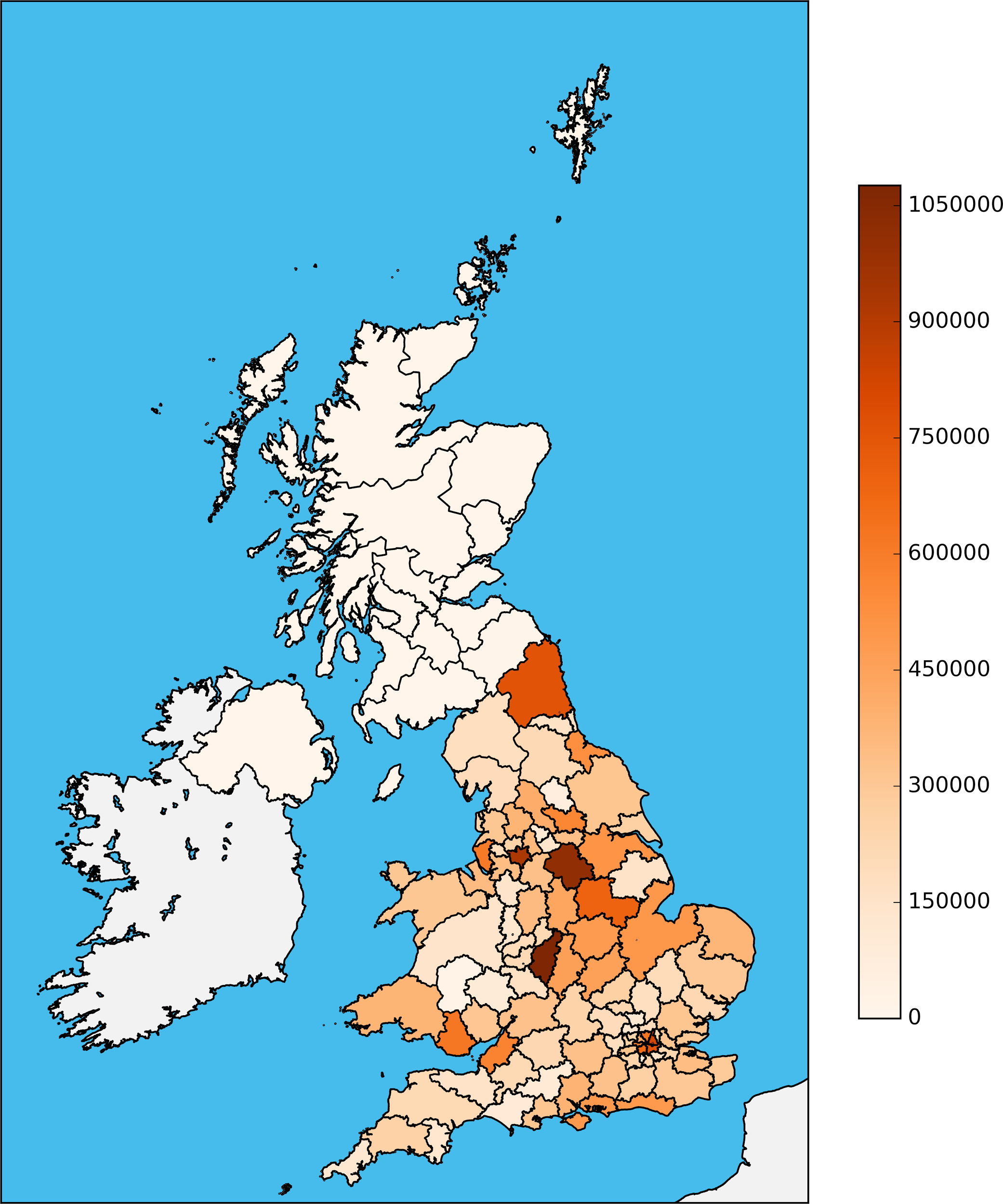
This section analyzes the distribution of the number of crimes in the UK at different granularities.

### 1.6.1. Total Number of Crimes in the Regions

The first approach was to plot the number of crimes by region. The darker the color of an area, the higher the number of crimes in that area. We can definitely see some areas with a higher number of crimes here. However the data should also be visualized at a lower granularity, as in [section 1.6.2](#_kzbfrvhktjm0).



### 1.6.2. Total Number of Crimes in the Postcode Areas

The map below depicts the number of crimes by postal area. The darker the color of an area, the higher the number of crimes in that area. Some darker areas can be observed which could be further analyzed to find the roots of the high number of crimes. 

1. http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-output-areas--soas-/index.html [↑](#footnote-ref-2)
2. http://www.pocketgpsworld.com/modules.php?name=POIs [↑](#footnote-ref-3)
3. https://data.gov.uk/dataset/national-statistics-postcode-lookup-uk/resource/3206f3b9-854a-46ec-8fd2-a6823e836b65 [↑](#footnote-ref-4)
4. https://www.police.uk/about-this-site/faqs/#what-do-the-crime-categories-mean [↑](#footnote-ref-5)
5. Used script: no. 2 [↑](#footnote-ref-6)
6. https://www.police.uk/about-this-site/faqs/#why-does-police.uk-not-show-outcomes-for-every-crime [↑](#footnote-ref-7)
7. https://www.gov.uk/government/uploads/system/uploads/attachment\_data/file/217421/police-uk-statistics.pdf [↑](#footnote-ref-8)