Compare Your Area

User Guide

October 2016

Contents

1. Introduction

The 'Compare Your Area' section of the police.uk website contains three charts that are designed to help you to answer the following questions:

How does crime in your local area compare with crime in other similar areas?

How does crime in your local area compare with crime in your police force?

How has crime changed over time in your local area and in your police force?

(What we refer to as a 'local area' is technically a 'Community Safety Partnership[1] area.)

In this context, two areas are said to be 'similar' if they have similar demographic, economic and social characteristics. The charts for your local area are generated automatically when you enter your postcode on entering the police.uk website.

The Compare Your Area charts were developed by a team of analysts in the Crime and Policing Analysis Unit of the Home Office.

To compare crime levels in your police force with crime levels in similar forces you can use a separate tool available on Her Majesty's Inspectorate of Constabulary's (HMIC's) Crime and Policing Comparator website. A link on the Compare Your Area page will take you directly to a chart for your police force on the HMIC website.

---

[1] There are currently 293 Community Safety Partnerships (CSPs) in England and 22 in Wales, the majority of which correspond to local authority areas. They are made up of representatives from the police, the local council, and the fire, health and probation services.

2. Data

Police recorded crime data

The Compare Your Area charts are based on published data on the numbers of crimes recorded by the police. Police forces return these data to the Home Office as part of their Annual Data Requirement and the data are published every three months by the Office for National Statistics (ONS). The data are designated 'National Statistics', which means that they have been produced in accordance with the Code of Practice for Official Statistics[2]. This requires the data to be produced, managed and disseminated to high professional standards.

The local area level crime data include the number of offences recorded under each offence type. When the data are processed to create the charts the individual offence types are aggregated into 14 crime types[3].

Population data

To calculate the rate of crime per thousand residents, mid-2015 population estimates from the ONS are currently used. These estimates are used to calculate all of the historical crime rates on the Compare Your Area charts. This approach ensures that crime rates for the current year are brought into line as quickly as possible with the basis on which crime statistics will be reported at the end of the year and avoids sudden changes of crime rates at the points where updated population estimates are applied. Note that the population basis of the published crime statistics is not generally updated after publication, so historical rates shown on the Compare Your Area charts will differ slightly from published rates in previous years. The mid-2015 population estimates are based on the 2011 Census.

---

[2] See the Code of Practice for National Statistics for more information: http://www.statisticsauthority.gov.uk/assessment/code-of-practice/index.html
[3] Offence types are aggregated into the following 14 crime categories: all crime, bicycle theft, burglary, criminal damage and arson, drugs, other theft, possession of weapons, public order, robbery, shoplifting, theft from the person vehicle crime, violence and sexual offences and other crime.

3. How to Interpret the Charts

Similar Local Area Bar Chart

This chart compares the crime rate in your local area to the average crime rate across similar areas. It shows the total number of crimes over a twelve month period per thousand residents, for the crime type selected.

Where your area lies in relation to the red and green lines is more important than its rank among similar areas. If your area lies between the red and green lines, its crime rate is normal for the group. If your area lies above the red line, its crime rate is higher than normal for the group, and similarly, if your area lies below the green line, its crime rate is lower than normal (See Appendix C for detail on how the red and green lines are calculated).

The areas shown in this chart are those that have been assessed to be most similar to your own. However, the circumstances within these areas do still vary and these variations can have an impact on the crime rates observed.

Within Force Bar Chart

This chart compares the crime rate in your local area to the average crime rate across the force area. It shows the total number of crimes over a twelve month period per thousand residents, for the crime type selected.

Trend Chart

This chart shows how crime rates in your local area and police force area have changed over time. It also shows how the average crime rates in similar areas to your local area have changed over time.

The chart shows the quarterly crimes per thousand residents, for the crime type selected, over a three year period. Note that some crime types have distinct seasonal patterns and this should be borne in mind when viewing the chart.

4. How Most Similar Groups are formed

What are Most Similar Groups?

Most Similar Groups (MSGs) are groups of local areas that have been found to be the most similar to each other using statistical methods, based on demographic, economic and social characteristics which relate to crime.

Areas which have similar demographic, economic and social characteristics will generally have reasonably comparable levels of crime.

Most Similar Groups are designed to help make fairer and more meaningful comparisons between areas. Community Safety Partnerships operate in very different environments and face different challenges. It is more meaningful to compare an area with the other areas which share similar socio-economic characteristics.
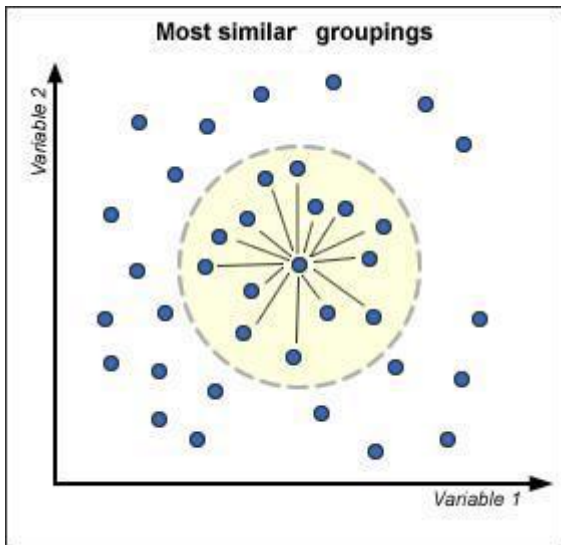
Generating Most Similar Groups

The development of the MSG approach involved stakeholders from the Home Office, Association of Chief Police Officers, Her Majesty's Inspectorate of Constabulary and other key stakeholders. The current approach was chosen following advice from independent academics.
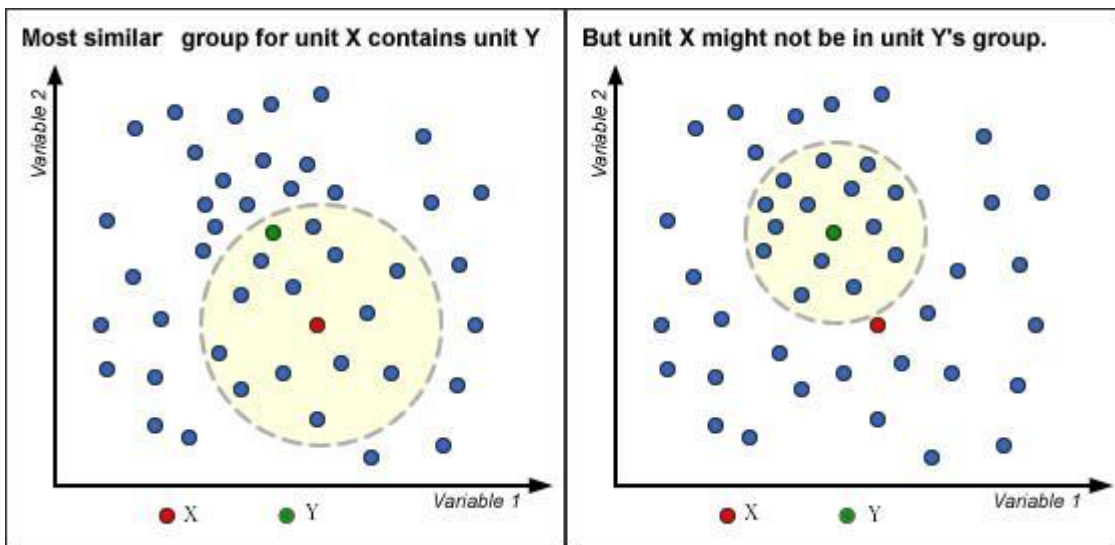
Twenty-four variables out of a possible 70 available were selected to generate MSGs (see Appendix A). They were identified by considering the levels of correlation with one or more of crime, fear of crime, or incidents. These variables are combined using a technique called Principal Component Analysis to determine new, uncorrelated variables that best describe the variation between areas. The so-called 'Most Similar Groups' are determined by identifying the areas which are most similar on the basis of these new variables.

Areas are compared in pairs to find the 'distance' between them for each variable. The overall distance between the pairs of areas is then calculated by summing the squared distances for all the variables. Each area is then grouped with up to 14 other areas to which it is 'closest', based on these distances. More technical detail on how these 'Most Similar Groups' are generated is provided in Appendix B.

The two-dimensional picture below shows an example identifying the 14 most similar areas to a given area based on two variables. This illustrates the method used to generate the Most Similar Groups.

Most similar groupings

Note that it is possible for area Y to be in area X's Most Similar Group without area X appearing in area Y's group. In the graphs below the 14 areas that are most similar to X are circled (left-hand side), one of which is area Y. However, X is not included in the 14 most similar areas to Y.

Most similar group for unit X contains unit Y

But unit X might not be in unit Y's group.

5. Appendix A

Variables used to calculate Most Similar Groups

Socio-demographic variables were chosen based on their correlation with crime levels. The full list of the 24 variables used to determine the local area groups is given below. They were chosen based on the levels of correlation with one or more of crime, fear of crime, or incidents.

1. **Percentage of ACORN 1 households.** ACORN is a proprietary (CACI) geodemographics dataset which assigns a neighbourhood description to each output area in the UK (the smallest geographical area at which Census data is available). ACORN 1 is referred to as " Wealthy Achievers ".
2. **Percentage of ACORN 2 households:** as above but for ACORN category 2 ("Urban Prosperity" neighbourhoods).
3. **Percentage of ACORN 4 households:** as above but for ACORN category 4 ("Moderate Means" neighbourhoods).
4. **Percentage of ACORN 5 households:** as above but for ACORN category 5 ("Hard Pressed" neighbourhoods).
5. **Percentage of student households.** The percentage of households categorised as student households from the 2011 Census.
6. **Percentage who have never worked.** The number of people who have never worked as a percentage of the 16-74 population from the 2011 Census.
7. **Percentage in routine/semi-routine occupations.** The number of people who are in routine or semi-routine occupations or have never worked as a percentage of the 16-74 population from the 2011 Census.
8. **Percentage permanently sick or disabled.** The percentage of people classified as permanently sick or disabled from the 2011 Census.
9. **Percentage of terraced households.** The number of terraced households divided by the total number of households (both from 2011 Census) multiplied by 100.
10. **Output Area (OA) density.** A population-weighted average of the density (population/area) of each OA. It aims to give a better indication of population density as it will highlight small pockets of densely populated housing.
11. **Percentage of overcrowded households.** From the 2011 Census. Households are classified as being overcrowded if they have an occupancy of more than 1 + number of bedrooms. This figure aims to represent the level of 'undesirable sharing' of rooms within a property.
12. **Percentage of single adult households.** The number of households containing only one person aged 18 or over (2011 Census) divided by the total number of households (2011 Census) multiplied by 100.
13. **Percentage of single parent households.** From the 2011 Census, the percentage of households which contain one parent and dependent children (15 and under, or 16-18 if in full-time education).
14. **Percentage of households with no working adults and dependent children.** From the 2011 Census, the percentage of households which contain dependent children (15 and under, or 16-18 if in full-time education) and no working adults.

15. **Population sparsity.** This variable gives an indication of the proportion of the population that lives in sparsely populated areas. It is equivalent to the sparsity measure used in the police funding formula.
16. **Long-term unemployed per worker.** From NOMIS, the number of people (average of July 2011 to June 2014) claiming job seekers allowance for more than 6 months, as a percentage of the population of working age.
17. **Long-term unemployed per claimant.** From NOMIS, the number of people (average of July 2011 to June 2014) claiming job seekers allowance for more than 6 months, as a percentage of total claimants.
18. **Percentage of 18-24 claimants.** From NOMIS, the number of people aged 18-24 (average of July 2011 to June 2014) claiming job seekers allowance, as a percentage of total claimants.
19. **Percentage of people on income support.** From NOMIS, the number of people (average of July 2011 to June 2014) claiming income support, as a percentage of the 16-74 population from the 2011 Census.
20. **Number of retail and leisure outlets.** This uses data supplied by a company called Retail Locations which collects location information on multiple retailers (i.e. chains / brands). The data are aggregated to calculate the number per hectare of retail and leisure outlets.
21. **Bars per hectare.** Uses data from the Annual Business Inquiry.
22. **Daytime population per hectare.** People who live and work in the area (or do not work) and those who live outside the area and work inside the area. It excludes those people who live in the area but work outside the area (2011 Census).
23. **Daytime net inflow (DTNI).** Change in the number of people in the area (either living or working) during the daytime (2011 Census).
24. **Percentage of population in hamlets or isolated dwellings.** The number of people living in hamlets or isolated dwellings as a percentage of the total population (2011 Census).

Appendix B

Generating Most Similar Groups – more detail

The Most Similar Group (MSG) generation involves four stages:

- Data preparation.
- Application of Principal Component Analysis to reduce the number of variables.
- Generation of the initial groupings.
- Pruning of the initial groupings to produce the final groups

The data preparation stage involves the following steps:

1. Calculate values of input variables for each area.
2. Transform the input variables by taking the natural logarithm.
3. Standardise the transformed variables by removing the mean and dividing by the standard deviation.

The Principal Component Analysis stage results in the number of variables being reduced from 24 to 4.

Generation of the initial groupings involves the following steps:

1. Calculate the 'distances' between all the areas, based on the variables generated by the Principal Component Analysis. (The distances used are Euclidean distances in 4-dimensional space.)
2. For each area, work out the 14 areas that are 'closest', based on the distances calculated in step 1.

Pruning of the initial groupings involves the following steps:

1. Calculate the standard deviation of the distance between each area and its median group member.
2. Calculate the distance between each area and the 'centroid' of its MSG, for each group member, if the group members were added sequentially.
3. Remove all group members for which the distance calculated in step 2 is larger than twice the quantity calculated in step 1.

Appendix C

Calculating the red and green lines

The calculation of the red and green lines (Upper and Lower bounds) involves 6 stages:

1. Calculate the 'normalised crime rate' for each area (CSP), defined as the crime rate divided by the MSG average.
2. Transform the normalised crime rates by taking the natural logarithm.
3. Calculate the mean and standard deviation of the transformed rates.
4. Use a standard result to convert this mean and standard deviation into the upper and lower quartile of the transformed rates.
5. Transform the upper and lower quartile back to the upper and lower quartile of the normalised rates by using the exponential function.
6. Multiply the resulting upper and lower quartile by the MSG average to obtain the upper and lower bounds.