CSEW Open Data tables

Introduction to Open Data tables

This chapter contains information on the content and use of the Crime Survey for England and Wales (CSEW) Open Data tables. These tables contain CSEW estimates related to victimisation and perceptions of crime and the criminal justice system (CJS) broken down by demographic characteristics. These tables are released as part of an initiative to make government data more transparent and accessible to the public and external researchers.

For users wanting to access full CSEW datasets, these are available from the [CSEW section](http://ukdataservice.ac.uk/get-data/key-data.aspx) of the [UK Data Archive](http://www.data-archive.ac.uk/) – please see chapter 10 for further information.

It is recommended that prior to using these Open Data tables, users read Chapter 2, in particular, of this user guide to familiarise themselves with the context of the data and the scope and limitations of the CSEW as a whole.

The Office for National Statistics (ONS) is now producing editions of the Open Data tables, following the responsibility for the publication of crime statistics transferring to ONS from the Home Office in April 2012. At present there are no plans to produce a back series to cover earlier CSEW years.

Table format

The main release consists of six data tables, all with an identical layout. The files have been split by theme, but because all have the same layout they can all be combined into a single table after download.

Tables are laid out in rows and columns with the first row containing column headings. Each row consists of a single CSEW estimate, while each column contains information about the estimate. Estimates are principally broken down by respondent characteristics (identified by the Sex, Age, HouseholdType, and Characteristic columns). However, other columns provide further information on the estimate, such as the time period it relates to (identified by the LastQuarter and LastYear columns). Further details can be found in the ‘Data table specification’ section of this chapter.

This layout is intended to provide CSEW estimates in the most transparent and versatile form for users and allow data to be linked with, for example, ONS population estimates for the calculation of numbers of crimes. Further details can be found in the ‘How to use CSEW Open Data tables’ section of this chapter.

All files are made available in CSV (comma separated values) format which arranges data in rows and columns as outlined above. This type of file can be opened in most data analysis programmes.

In addition to the six main tables, two additional reference tables have been included in this release. See the ‘Reference data tables’ section of this chapter for more information on these.

File naming

The six main tables are contained in files which use the following naming structure:

‘[Measurement Type]-[Measurement Subcategory]-[Date].csv’

Eg Household-Incidence-2014-Q4.csv

The [Measurement Type] and [Measurement Subcategory] sections of the filename are used to identify the content of the data file. There are three different values for [Measurement Type], each containing two values for [Measurement Subcategory]. An outline of what these values are and a description of what is contained in the files can be found in Table 9a.

The [Date] section of the filename identifies the CSEW dataset that the release was taken from. It follows the same convention as the LastYear and LastQuarter columns in the tables themselves (see the ‘Data table specification’ section of this chapter).

|  |  |  |  |
| --- | --- | --- | --- |
| Measurement type | Measurement subcategory | Description | Example of data in file |
| Household | Incidence | Contains incidence rates for household crime | Number of incidents of burglary per 1,000 households in the 12 months prior to interview |
| Prevalence | Contains prevalence rates for household crime | Percentage of households that have been victims of burglary in the 12 months prior to interview |
| Personal | Incidence | Contains incidence rates for personal crime | Number of incidents of violence per 1,000 adults in the 12 months prior to interview |
| Prevalence | Contains prevalence rates for personal crime | Percentage of adults who have been victims of violence in the 12 months prior to interview |
| Perceptions | CJS | Contains perceptions of the police and the criminal justice system | Percentage of adults who think that their local police are doing a good or excellent job |
| Other | Contains other perceptions of crime | Percentage of adults who perceive a high level of anti-social behaviour in their local area |

Data table specification

Data tables are laid out in rows and columns. Each row contains a single CSEW estimate, which each column contains information about that estimate. Below is a description of the meaning behind the values in each column.

SOURCE

Possible values: CSEW

This column identifies the source of the estimate. For this release, this column contains only one value: ‘CSEW’. This field is provided to allow future releases to include data from different sources.

LASTYEAR

Possible values: Various

Combined with the LastQuarter column, this column identifies the CSEW interview period that the estimate is based on. The combination of LastYear and LastQuarter identifies the last quarter of interviews covered. For example, **LastYear** = ‘2014’ and **LastQuarter** = ‘4’ denotes that the estimate is based on the 12 months of CSEW interviews between January 2014 and December 2014 – i.e. the 12 months ending in quarter 4 of 2014.

LASTQUARTER

Possible values: Various

Combined with the LastYear column, this column identifies the CSEW interview period that the estimate is based on. The combination of LastYear and LastQuarter identifies the last quarter of interviews covered. For example, **LastYear** = ‘2014’ and **LastQuarter** = ‘4’ denotes that the estimate is based on the 12 months of CSEW interviews between January 2014 and December 2014 – i.e. the 12 months ending in quarter 4 of 2014.

MEASUREMENTVAR

Possible values: Various (see reference table)

This is the name of the variable that is being measured by the estimate. It could also be considered a dependent variable. In this release it is a crime incidence rate, a crime prevalence rate or a perception of crime. Future releases may contain a wider range of measurements. The variable name is a shorthand way of referring to the measurement in question, therefore the meaning of the values in this column may not be immediately obvious to users. A reference table has been provided with this release that contains more descriptive labels for each of the variable names in this column. For users who have access to the main CSEW dataset from the UK Data Archive, the names here match directly with the variable names on the main CSEW dataset.

GEOGRAPHY

Possible values: England and Wales

This column identifies the geographical area that the estimate relates to. For this release, only estimates for the whole of England and Wales are included. Future releases may provide regional breakdowns.

**AGE**

**Possible values: 16+; 16-24; 25-34; 35-44; 45-54; 55-64; 65-74; 75+**

This column identifies the age of the respondent at the time of interview. Respondents are arranged into banded age groups, with the category of ‘16+’ covering respondents in all age groups combined. This column only contains values where MeasurementLevel is ‘Person’ (see reference table).

**SEX**

**Possible values: All adults; Male; Female**

The column identifies the sex of the respondent. The category of ‘All adults’ covers men and women combined. This column only contains values where MeasurementLevel is ‘Person’ (see reference table).

HOUSEHOLDTYPE

Possible values: All households; Vehicle-owning households; Bicycle-owning households

This column identifies the type of household that the estimate relates to, specifically whether the household owns a vehicle or bicycle. Most estimates relate to ‘All households’. ‘Vehicle-owning households’ and ‘Bicycle-owning households’ provide an alternative measure for the prevalence of vehicle-related crime and bicycle theft respectively and as such only have a value for these measures. This column only contains values where MeasurementLevel is ‘Household’ (see reference table).

CHARACTERISTICVAR

Possible values: Various (see reference table), GOR, Total

This column identifies the variable that is used to provide the full breakdown of estimates by respondent characteristics in the Characteristic column. It could also be considered an independent variable. It is similar to the MeasurementVar column in that it contains variable names that match the variable names on the main CSEW dataset. More descriptive labels for the variable names in this column can be found in the reference table for this column.

The value ‘GOR’ identifies estimates for each of the English regions and for England and Wales separately. The value ‘Total’ in this column identifies an estimate that covers all respondents, regardless of individual characteristics. Users should be aware that some characteristics only exist at the personal level and some only exist at the household level (identified in the MeasurementLevel column – see reference table).

CHARACTERISTIC

Possible values: Various; Total

This column identifies respondent characteristics that the estimate relates to. Characteristics are grouped by CharacteristicVar such that each value of CharacteristicVar has a number of unique values for Characteristic associated with it. As well as those characteristics included on the main CSEW dataset within each CharacteristicVar, the Open Data tables also include some combined categories that are not included on the CSEW core variables. The value ‘Total’ in this column identifies an estimate that covers all respondents, regardless of individual characteristics and is the only Characteristic for the CharacteristicVar of ‘Total’.

**ESTIMATE**

**Possible values: Various**

This column contains the estimate for MeasurementVar for respondents with the characteristics referred to at Geography, Age, Sex, HouseholdType and Characteristic from CSEW interviews conducted in the period identified by LastQuarter and LastYear. When MeasurementType is ‘Incidence rate’ it is a rate per 1,000 adults/households; where MeasurementType is ‘Prevalence rate’ or ‘Perception’ it is a percentage (see reference table).

STANDARDERROR

Possible values: Various

This column contains the standard error of the value of Estimate. As the CSEW is a sample survey, all estimates are subject to a degree of error, reflected in the size of the standard error – see the ‘How to use CSEW Open Data tables’ section below for more information. More advanced users may also want to know that as the CSEW is based on a complex sample design, the standard errors included in these tables are complex standard errors that take the sample design into account. This means that no further adjustments, such as design factors, need to be applied to calculations involving these standard errors.

UNWEIGHTEDCOUNT

Possible values: Various

Also referred to as the ‘Unweighted base’, this column shows the total number of respondents who contributed to the calculation of Estimate. Estimates based on a larger number of respondents are generally more reliable, and those based on a very small number of respondents should be treated with caution. It is recommended that users refrain from using estimates based on fewer than 50 respondents. They have been included here for reference and in the interests of completeness.

Reference data tables

There are two reference data tables included with this release; one each for the columns MeasurementVar and CharacteristicVar. These provide additional information on the variables included in these two columns in the main data tables. They can be used as lookups to include this additional data in combined tables. The data specification for these tables is below.

Measurement

MEASUREMENTVAR

Possible values: Various

This includes all the values in the MeasurementVar column of the main data tables.

MEASUREMENTLABEL

Possible values: Various

This column contains a text description of the measure that MeasurementVar relates to.

MEASUREMENTLEVEL

Possible values: Person; Household

This column identifies the level at which the measurement applies. For personal crimes and perception measures, this column has a value of ‘Person’ to indicate that the estimate applies to people (specifically adults aged 16 and over). For household crimes, this column has a value of ‘Household’ to indicate that the estimate applies to households in England and Wales.

MEASUREMENTTYPE

Possible values: Incidence rate; Prevalence rate; Perception

This column identifies the type of measurement. ‘Incidence rate’ means that the measurement shows the number of crimes per 1,000 adults or per 1,000 households (identified by the MeasurementLevel column) in the last 12 months. ‘Prevalence rate’ means that the measurement shows the percentage of adults/households (identified by the MeasurementLevel column) who have been victims of crime in the last 12 months. ‘Perception’ means that the measurement shows the percentage of adults who have a particular perception about crime or the criminal justice system.

Characteristic

CHARACTERISTICVAR

Possible values: Various

This includes all the values in the CharacteristicVar column of the main data tables.

CHARACTERISTICLABEL

Possible values: Various

This column contains a text description of the measure that MeasurementVar relates to.

How to use Open Data tables

At the simplest level, CSEW Open Data tables can be used to find CSEW estimates for certain demographic groups. The CSV files can be imported into most data analysis programmes for this purpose. A full list of the measurements that are included in these tables can be found in the **MeasurementVar** reference table. A full list of demographic characteristics that these measurements can be analysed by using these tables can be found in the **CharacteristicVar** reference table.

The data specification table above should make clear to users that to find the estimate they are interested in requires filtering or searching data across different columns to identify the population group they are interested in. For example, overall national estimates for all adults aged 16 and over can be found by selecting ‘16+’ from the Age column, ‘All’ from the Sex column, and ‘Total’ from the Characteristic column (for personal crime or perception estimates). At the other extreme level of detail, users could find estimates for 16-24 year old married women by selecting the appropriate categories from the Age, Sex and Characteristic columns. Other levels of detail can be found by selecting the appropriate values in these fields (or HouseholdType for household crime estimates).

When using these estimates, users should be aware of the fact that the CSEW is a sample survey and produces estimates with a margin of error around them. As such, ONS recommends that users do not use estimates based on fewer than 50 respondents (identified by the UnweightedCount column). This feature of the data should also be considered when comparing differences between groups. Standard statistical tests can be used to identify whether differences between demographic groups are ‘statistically significant’. That is, whether the differences seen in the CSEW data are due to differences in the population of England and Wales as a whole, or whether they occurred by chance as a result of the random selection of respondents for the CSEW sample. Users should familiarise themselves with this concept before using these tables.

When comparing estimates, users should ensure that they are only comparing estimates from independent samples. Although the current set of tables only include data for a single time period, future releases may cover other time periods. When comparing data over time, users should consider the time periods that are being used to ensure that they are not overlapping. This can be identified by consideration of the LastYear and LastQuarter columns in the data tables.

Some users may want to combine data in these tables with data from other sources. One common use of this would be to produce counts of crime and numbers of victims by combining the CSEW Open Data tables with population data. Incidence rates in these tables show the number of crimes per 1,000 adults or 1,000 households. By multiplying the values in the CSEW Open Data tables by ONS estimates of the total number of adults/households in England and Wales divided by 1,000, it is possible to calculate the total number of crimes that took place over a 12 month period. Similarly, prevalence rates show the percentage of adults or households who were victims of crime in a 12 month period. If these percentages (once converted to a decimal) are multiplied by the total number of adults/households in England and Wales, it is possible to calculate the number of victims of crime. Users should bear in mind that when doing this, it is important to use a population estimate for the same subpopulation as that used in the CSEW Open Data tables. So, for example, an incidence rate for men aged 16-24 in England and Wales from the CSEW Open Data tables should be multiplied by a population estimate for men aged 16-24 in England and Wales to calculate the number of crimes against men aged 16-24 in England and Wales.

Changes to the Open Data tables variables

For the survey year 2012/13 (i.e. interviews from April 2012 onwards) changes were made to the demographic questions in the CSEW questionnaire. The changes have been made in accordance with harmonised standards set by ONS with the aim of providing consistent data series across government ([ONS Harmonisation)](http://www.ons.gov.uk/ons/guide-method/harmonisation/harmonisation-programme/index.html). As these questions were introduced into the survey in April 2012, any annual dataset produced over the previous 12 month period will necessarily have included both the new and the old style questions. This affected datasets for the reporting periods ‘Year to June 2012’, ‘Year to September 2012’ and ‘Year to December 2012’.