

NCRB Data Notes

Unit of area in the NCRB

In the original data provided by NCRB, the unit of observation does not exactly correspond to the districts in India. Although a majority of the districts do feature; additional units such as the railways of a certain district, rural areas of certain districts, or even cities within a particular district in the state are found in different instances, across different years. To cite a few examples, Guntakal Railways (Guntakal itself comes under the Anantapur district), Vijayawada City, Vijayawada Railway, Special Task Force (S.T.F), Anti Terrorist Squad etc. As a result, the number of these places are much more than the number of districts in the country, and also happen to vary each year. One of the steps in cleaning the data was trying to ensure that each of these places were mapped to the districts since we show the values at a district level.

| Place recorded in NCRB dataset | District it was mapped to |
|--------------------------------|---------------------------|
| Jaipur East | Jaipur |
| Jaipur West | Jaipur |
| Bettiah | Pashchim Champaran |
| Motihari | Purba Champaran |
| Hansi | Hisar |
| Baddi | Solan |

For instances which are similar to the first two rows in the above table, we used a combination of fuzzy matching and regex expressions in order to categorize them into the corresponding districts. However for other variants, it is also a matter of manually searching for the cities / areas within a district and then adding them to the crimes of that particular district (Hisar and Solan also appear in the NCRB data, and so do Handi and Baddi, their respective cities which were then mapped to their parent districts).

Similarly a combination of code and manual inspection was used to address the inconsistencies in the spellings. Like mapping *cbpura* to Chikkabalapura, *sbs nagar* to Shahid Bhagat Singh Nagar or to differentiate districts like Sangareddy and Rangareddy (where the algorithms would have classified them as the same district). In some cases, places / districts were known by more than one name / or their names were changed after a certain

point in time and as a result were inconsistent. A few of these examples are tabulated in the table below

| Place recorded in NCRB Data | Name in the Census data |
|-----------------------------|-------------------------|
| Prayagraj (after 2018) | Allahabad |
| NC Hills | Dima Hasao |
| Gurugram (after 2016) | Gurgaon |
| Hathras | Mahamaya Nagar |

Categories of crime in the NCRB

From the year 2001 until 2013, the crime categories provided by the NCRB are lesser and also less detailed than the categories provided from the year 2014 onwards. From 2014, NCRB is reporting more IPC crime sections (and so we have more disaggregated crime categories). For example if only the incidences of riots were reported in the earlier time period (before 2014), more granularity or the types of riots were also reported in the year 2014 onwards, such as communal / religious, sectarian, industrial, political, caste conflict etc. This is the case with almost every other category. Hence, we have two series on the CEDA data portal, an [IPC crime series](#) and a [comprehensive crime series](#). Since there are more sub categories for data available from 2014 onwards, they have been summed up to form the total categories of the data available from the period 2001 to 2013. This way, we obtain a series which has the data for the entire period. In the second instance, since we have granular data for the period 2014 onwards, the comprehensive series consists of the various subcategories which have been provided by NCRB.

Crime Incidence and Crime Rates

The user has the option to download the crime incidences and the crime rates (cases per lakh population). For calculating the crime rates, we used the population data from Census and used interpolation between the two Census years, assuming a constant growth rate. To corroborate values at times, we also compared the difference in the values from the method obtained above and the population

projections at the state level as found in the Census estimates found [here](#). In case the user wants to use their own calculated values in order to obtain the crime rates, they can do so by just obtaining the number of incidences from the portal.