

Uniform Crime Reporting: Summary Reporting System Datasheet

I. MOTIVATION

I-A For what purpose was the dataset created?

The Summary Reporting System (SRS) is part of the FBI's Uniform Crime Reporting program. SRS aims to profile a picture of crime in the United States by collecting monthly agency-level counts of reported criminal offenses. These counts are provided by participating police agencies.

I-B Who created the dataset?

Is it an official law enforcement or government body? An academic research team? Other?

The dataset is provided by participating law enforcement agencies, and compiled as part of the FBI's Uniform Crime Reporting program.

I-C Was there a specific task in mind?

Outside of collecting aggregated crime statistics there is no specific task in mind.

I-D Was there a specific task in mind, or gap that needed to be filled?

The uniform crime reporting program was established in 1930 to serve as a periodic nationwide assessment of reported crimes, which were not available elsewhere in the criminal justice system.

II. COMPOSITION

II-A What do the instances that comprise the dataset represent?

For example: crimes, offenders, court cases, police officers

The SRS reports on arrests at an aggregate level. Each instance corresponds to the number of arrests, conditional on crime and demographics of the offender, from an individual reporting agency.

II-B Are there multiple types of instances?

For example: offenders, victims, and the relationship between them.

No.

II-C How many instances are there in total?

Of each type, if appropriate.

In 2020, the UCR SRS program includes data collated from more 15,875 city, university and college, county, state, tribal, and federal law enforcement agencies out of a total 18,623 agencies, covering 98% of the U.S. population.

II-D Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?

For example, if it is traffic stops from a territory, is it all traffic stops conducted within that territory within a specific time? If not, is it a representative sample of all stops? Describe how representativeness was validated/verified. If it is not representative, please describe why.

Agency participation is voluntary, as such, SRS does not contain all agencies within the United States. The arrests reported as all the arrests from the reporting agency.

II-E What data does each instance consist of?

If there is a large number of variables, please provide a broad description of what is included.

Each instance contains counts the number arrests for Part I offenses:

- Homicide
- Rape
- Aggravated Assault
- Robbery
- Burglary
- Larceny-theft
- Motor-vehicle theft
- Arson
- Human Trafficking

and for Part II offenses:

- simple assault
- curfew offenses and loitering
- embezzlement
- forgery and counterfeiting
- disorderly conduct
- driving under the influence
- drug offenses
- fraud
- gambling
- liquor offenses
- offenses against the family prostitution, public drunkenness, runaways, sex offenses, stolen property, vandalism, vagrancy, and weapons offenses

Additionally, information is collected on demographics, such that one can condition on race, age, or sex to get the arrest count for that demographic.

II-F Is there a target label or associated with each instance?

Please include labels that are likely to be used as target labels, e.g. recidivism.

No.

II-G Are there recommended data splits (e.g., training, development/validation, testing)?

If so, please provide a description of these splits, explaining the rationale behind them.

No.

II-H Does the dataset contain data on race and ethnicity?

If so, is it based on the individual's self-description, or based on officer's impression? Was it collected or derived in post-processing? For example, by name analysis.

The dataset contains data on race only. This is derived from the officer's impression, unless they choose to ask the arrestee about their race.

II-I Are there any known errors, sources of noise, bias or missing data, or variables collected for only part of the datasets?

If so, please provide a description.

No.

II-J Does the dataset contain data on criminal history or other data that might be considered confidential or sensitive in any way?

For example: sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; If so, please provide a description.

No. Data is not on an individual level.

II-K Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset?

If so, please describe how.

No.

III. USES

III-A What type of tasks, if any, has the dataset been used for?

If so, please provide examples and include citations.

The dataset has been used for many tasks, a non-exhaustive repository can be found:

<https://www.icpsr.umich.edu/web/ICPSR/series/57>

III-B Is there a repository that links to any or all papers or systems that use the dataset?

If so, please provide a link or other access point.

Please see above.

III-C What (other) tasks could the dataset be used for?

For example: testing predictive policing systems, predicting recidivism.

The dataset can be used for tasks which needs a count of the number of arrests, at an agency level, reported throughout the United States. This dataset is most useful when used in conjunction with other datasets.

III-D Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?

For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (e.g., stereotyping, quality of service issues) or other risks or harms (e.g., legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?

The SRS uses a "hierachy rule" when counting reported offending. This means only the *worst* offense associated with an arrest is counted. It is also important to note that this only includes arrests, not all crime that occurs.

IV. COLLECTION PROCESS

IV-A How was the data associated with each instance acquired?

e.g. the data collected survey, the raw data is routinely collected by the courts.

Data is submitted by a participating law enforcement agencies.

IV-B Was the information self-reported?

If the data was self-reported, was the data validated/verified? If so, please describe how.

No. The data is not self reported. The raw data comes from law enforcement agencies day-to-day data collection.

IV-C Who was involved in the data collection process?

Was this done as part of their other duties? If not, were they compensated?

Participating police agencies. Raw data is collected routinely as part of policing work.

IV-D Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)?

If not, please describe the timeframe in which the data associated with the instances was created. If the collection was not continuous within the timeframe, please specify the intervals, for example, annually, every 4 years, irregularly.

The data has been collected since 1930. Annual data releases from 1980 onwards are available on the UCR website.

IV-E Were any ethical review processes conducted (e.g., by an institutional review board)?

If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.

Unknown (unlikely).

IV-F Were the individuals in question notified about the data collection? Did they give their consent?

If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses?

No. Consent is not granted as the individuals have no option to opt-out.

IV-G Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis) been conducted?

If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.

Unknown (unlikely).

V. PRE-PROCESSING, CLEANING, LABELING

V-A Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, removal of instances, processing of missing values)?

If so, please provide a description and reference to the documentation. If not, you may skip the remaining questions in this section.

Incidents are classified according to the hierarchy rule by each agency before counting. For further detail view the documentation [1].

V-B Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data?

If so, please provide a link or other access point to the “raw” data.

The raw data is not available.

V-C Is the software that was used to preprocess/clean/label the data available?

If so, please provide a link or other access point.

N/A

VI. DISTRIBUTION

VI-A Is the data publicly available? How and where can it be accessed (e.g., website, GitHub)?

Does the dataset have a digital object identifier (DOI)?

Yes. The data is available to download from the <https://crime-data-explorer.fr.cloud.gov/pages/home> website.

VI-B Is the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?

If so, please describe this license and/or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.

The dataset is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) License.

VII. MAINTENANCE

VII-A Is the dataset maintained? Who is supporting/hosting/maintaining the dataset?

The FBI UCR program.

VII-B How can the owner/curator/manager of the dataset be contacted (e.g., email address)?

The owners can be contacted at: UCR-SRS@fbi.gov

VII-C Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)?

New data is released annually.

VII-D Are older versions of the dataset continue to be supported/hosted/maintained?

Yes. Data releases dating back until 1980 are hosted on the UCR website.

VII-E If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?

If so, please provide a description.

No.

REFERENCES

[1] “Summary Reporting System User Manual.”