

Datasheet for School Shootings Dataset*

Data on School Shootings in the U.S. Since Columbine

Yun Chu

Based on Washington Post Reporting

2024-12-03

This datasheet provides information on a dataset documenting acts of gunfire at primary and secondary schools in the U.S. since Columbine. It outlines the dataset's motivation, composition, collection process, and potential uses. It aims to support informed analysis of school shootings and their impacts.

Extract of the questions from Gebru et al. (2021).

1 Datasheet for School Shootings Dataset

1.1 Motivation

1. Purpose:

- The dataset was created to identify every act of gunfire at a primary or secondary school during school hours since the Columbine High massacre on April 20, 1999. This dataset aims to quantify the impact of school shootings on children.

2. Creators:

- The Washington Post team, including researchers and reporters such as John Woodrow Cox, Steven Rich, Allyson Chiu, and others.

3. Funding:

- The Washington Post.

4. Comments:

- The dataset was curated using Nexis, news articles, open-source databases, law enforcement reports, school websites, and direct communication with schools and police departments.

*Code and data are available at: <https://github.com/washingtonpost/data-school-shootings>

1.2 Composition

1. Instances Represented:

- Each instance represents an act of gunfire at a U.S. primary or secondary school during school hours.

2. Total Instances:

- The dataset contains over 416 incidents of gun violence that fit The Post’s criteria.

3. Completeness:

- This dataset is a comprehensive attempt but may not include every qualifying incident due to the lack of federal tracking.

4. Data Content:

- Raw data includes variables such as date, location, injuries, casualties, and school demographic information.

5. Labels or Targets:

- No specific target variable, but the dataset includes classifications of incidents.

6. Missing Information:

- Some details may be unavailable or redacted.

7. Relationships Between Instances:

- No explicit relationships beyond shared metadata like location or date.

8. Data Splits:

- Not applicable.

9. Errors or Noise:

- Efforts were made to validate data through cross-referencing.

10. External Resources:

- Enrollment figures and demographic data from the U.S. Department of Education.

11. Confidentiality:

- No individual-level confidential information is present.

12. Sensitive Content:

- The dataset contains data on gun violence, which may be distressing.

13. Subpopulations:

- Identified by school demographics.

14. Identification of Individuals:

- Individuals cannot be directly identified.

15. Sensitive Data:

- Includes data on school violence and student demographics.

1.3 Collection Process

1. Acquisition Method:

- Data was gathered from public records, news articles, and direct verification with schools and law enforcement.

2. Mechanisms:

- Manual curation and data validation.

3. Sampling Strategy:

- Incidents were selected based on predefined criteria.

4. Contributors:

- Washington Post reporters and researchers.

5. Timeframe:

- April 20, 1999, to Sep 6th, 2024.

6. Ethical Review:

- No formal review disclosed.

7. Source of Data:

- Third-party sources such as schools, law enforcement, and public databases.

8. Notification to Individuals:

- Not applicable.

9. Consent:

- Not applicable.

10. Revocation Mechanism:

- Not applicable.

11. **Impact Analysis:**

- Not disclosed.

1.4 **Preprocessing/Cleaning/Labeling**

1. **Preprocessing:**

- Incidents not meeting criteria (e.g., after-hours events, accidental discharges) were excluded.

2. **Raw Data Retention:**

- Raw data is accessible via GitHub.

3. **Software Used:**

- Details not provided.

1.5 **Uses**

1. **Prior Uses:**

- Analyzed by The Washington Post for reporting and visualization.

2. **Repository:**

- Available on [GitHub](#).

3. **Potential Tasks:**

- Statistical analysis, policy research, educational interventions.

4. **Impact Considerations:**

- Users should handle data with sensitivity and avoid unintended biases.

5. **Prohibited Uses:**

- Using data to sensationalize or misrepresent school violence.

1.6 Distribution

1. Third-Party Sharing:

- Shared publicly via GitHub.

2. Distribution Methods:

- GitHub repository.

3. Availability Timeline:

- Dataset is continuously available.

4. Licenses:

- This data is published under an Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) license. More details are available [here](#).

1.7 Maintenance

1. Support:

- Maintained by The Washington Post.

2. Contact:

- schoolshootings@washpost.com.

3. Erratum:

- Not specified.

4. Updates:

- Updates are dependent on new incidents and criteria.

5. Retention Limits:

- Not applicable.

6. Versioning:

- Versioning policies not specified.

7. Contributions:

- Not applicable.

References

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2021. “Datasheets for Datasets.” *Communications of the ACM* 64 (12): 86–92. <https://doi.org/10.1145/3458723>.