

# Data

## Data

The main analysis utilizes Daily Crime Logs from each university's dedicated police department over a six year period (2014-2019). I omit the year 2020 due to the COVID-19 pandemic in which university activity varied substantially from the previous years. Under the Jeanne Clery Act, Universities that receive federal funding are mandated to keep Daily Crime Logs which specify the universe of crimes that university police officers report or are reported to the university police department over the last seven years. Daily Crime Logs are unique in that the data is unaggregated; each Daily Crime Log contains the date reported at the hourly level, in addition to a short description of the crime. Moreover, the Daily Crime Logs contain all incidents reported by or to the university police. Therefore, the data includes offenses such as alcohol offenses which are typically missing from national databases such as the Uniform Crime Reporting System (UCR) and the National Incidence-Based Reporting System (NIBRS). Additionally, the by-hour reporting makes the Daily Crime Logs preferable to other university crime databases such as Campus Safety and Security Data provided by the US Department of Education which features only yearly-level data on offenses.

## Data Collection

Under the Jeanne Clery Act, university police must allow their Daily Crime Logs to be ready for inspection within two business days of an inquiry. The universities will then either send their Daily Crime Logs (usually in PDF format) through email, direct you to a website that holds their records, or arrange to set a time to inspect records.<sup>1</sup> Each of these records were parsed using PDF extracting and webscraping techniques. Figure **SAMPLE OF CRIME LOG** shows a sample of a Daily Crime Log.<sup>2</sup>

## Data Harmonization

Each university's Daily Crime Log varies in the description of the crimes reported. To achieve harmonization across university police departments, I pattern-matched key words using regular expressions relating to specific categories of crimes that were taken from the US Department of Education Campus Safety and Security Data: sexual assault, alcohol offenses, drug offenses, theft, and robbery. Table **KEYWORDS** shows the keywords used to pattern match to each offense type. These keywords were derived from surveying the most frequent descriptions of crimes within each university police department. For instance, if the phrase "liquor law violation" was a common description in a university's crime logs, then the keyword "liquor" was used to match to an alcohol violation. While this method is imperfect—it may underestimate the true reported offenses if particular keywords were not matched or it may overestimate the true reported offenses if particular keywords systematically match to incorrect crimes—Table **MOST COMMON WORDS** shows the 15 most frequently reported offenses after the pattern matching process. In each of the columns, there is no apparent mismatching. Furthermore, Figure **VENNDIAGRAM** shows that there is little overlap between the three main outcomes of alcohol, drug, and sexual assault offenses.

---

<sup>1</sup>Universities did not have to allocate copies of the crime logs, they only had to make them available for inspection. Hence, there are **SEVERAL UNIVERSITIES** that may have had fraternity moratoriums, but were not able to be included in the sample. Additionally, **NUMBER OF UNIVERSITIES WITH UNREADABLE** universities provided Daily Crime Logs that were in formats that were completely unreadable by any software (e.g scanned documents). These **BLANK NUMBER** were

<sup>2</sup>Nearly every Daily Crime Log varied by university.