A. The Dataset

The dataset is from National Crime Victimization Survey (NCVS) collected from January 1, 2020, to December 31, 2020, about personal and household victimization in the US. There are four objectives in NCVS: to develop detailed information about the victims and consequences of crime, to estimate the numbers and types of crimes not reported to the police, to provide uniform measures of selected types of crimes, and to permit comparisons over time and types of areas (ICPSR 38090, pg.4). This data set is not from an experiment rather an ongoing survey starting in 1972. NCVS is classified as prospective study because it's not based on historical data but rather aims to capture data as events unfold through scheduled surveys and interviews.

There are 81 variables in this data set (age, race, location of crime, etc.) and 8043 observations. Overall, there are 1811 distinct households in the dataset. The NCVS uses 'rotating panel' sample design. Randomly chosen households are selected to be part of the NCVS panel. All age-eligible individuals (at least 12 years old) in these households become panel members. Panel members are interviewed every six months for a total of seven interviews over three years. After the seventh interview, the household leaves the panel. A new household is rotated into the sample to replace the departing household. It is an unbiased sampling because the survey uses a stratified, multistage cluster sampling design (ICPSR 38090, pg.9). This technique is designed to reduce potential biases and produce a sample that represent various geographic and demographic groups in the US.

B. Characteristics of Sample

The NCVS 2020 dataset has a lot of categorical variables like employment, education, marital status, etc. Out of 81 variables in the dataset, only 15 of them are numeric variables such as the number of reports, number of times of being stolen, and age. Before diving into the relationship between variables, we will first examine several variable distributions in the dataset.

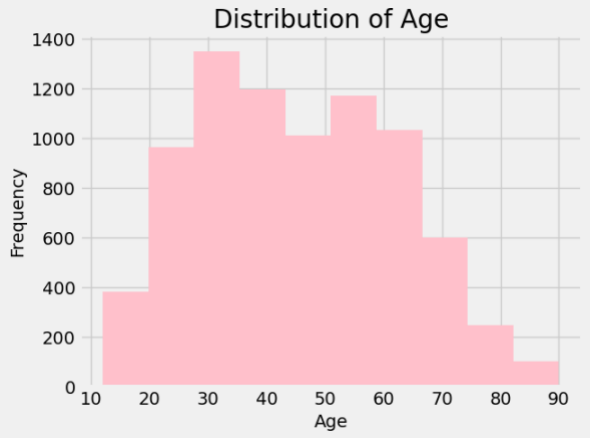

Figure 1. Age Distribution

In the histogram on the left (Figure 1), we get the age distribution of 2020 NCVS. It's evident that most respondents fall within the 30-40 age range, with the frequency around 1300, being the highest bin in the histogram. Since only people above 12 years are only eligible to be in NCVS, the histogram starts at age equals 12 and ends around age interval 80-90. Additionally, the histogram is right skewed which means the dataset has a higher proportion of older individuals than the younger ones. Some statistics about the histogram: the mean age is 45.24 years, the median is 44 years, and the standard deviation is 16.92 years.

	Marital Status	Count
0	Married	3131
1	Never married	2760
2	Divorced	1333
3	Widowed	475
4	Separated	301
5	Residue	43

Table 1. Marital Status

The table on the left (Table 1) shows the marital distribution of respondents in the NCVS 2020 survey. The biggest number in the data is 3,131, which represents the count of people who said they are "Married." On the other hand, the smallest number is 43, and it represents the "Residue" category. This information tells us that the most common marital status among the people surveyed is "Married," showing that a lot of people in the survey fit into this group. The data also shows the different marital statuses in the dataset, including "Never married," "Divorced," "Widowed," and "Separated," which fall between the highest and lowest values.

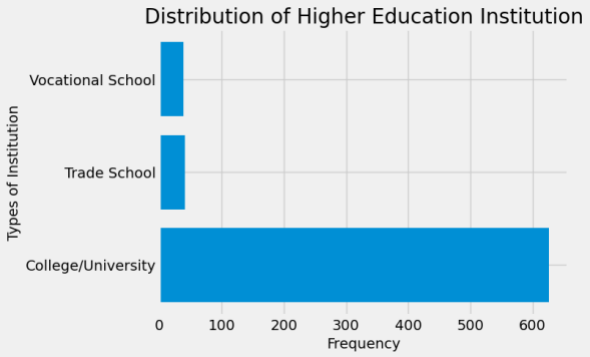

Figure 2. Higher Education Distribution

Figure 2 focuses on students who are attending higher education institution (above high school). Within the context of higher education, it becomes clear that most higher education respondents in this survey have opted for the traditional route, enrolling in colleges and universities. This choice is particularly dominant, which approximately 600 more students than any other institution. In contrast, the data reveals two distinct outliers, representing vocational schools and trade schools, with both options drawing frequencies of fewer than 50 students each.

C. Relationship Between Variables

This report will further analyse variables related to college life and criminal incidents. The objective of this report is to unravel the correlations between these variables and find out how factors such as students' types of housing influence both the frequency of crimes they experience and their likelihood of reporting crime to the police.

	where students live	number of students	total_number_incidents	mean
0	off-campus	663	958.0	1.444947
1	on-campus	43	38.0	0.883721

Table 2. Mean Crime per Student

Table 2 above presents data about the living arrangements of college students and their corresponding experiences with incidents or crimes. The information is categorized into two groups based on where they live: "off-campus" and "on-campus." This table offers insights of the relationship between where students live and their exposure to incidents or crimes. Notably, majority of students (633) reside off-campus, while a smaller group (43) lives on-campus.

Off-campus students have reported a significantly higher total number of incidents (958) compared to on-campus students (38). The average of incidents per student for this group is approximately 1.44 which is again relatively higher than on-campus students which only has a mean of 0.88. This indicates, on average, on-campus experience fewer incidents.

	Student Type	Mean Break-ins	Mean Stolen	Mean Vehicle Thefts	Mean Crimes with Weapon	Mean Forced Sex
0	On Campus	0.023256	0.697674	0.000000	0.093023	0.069767
1	Off Campus	0.119155	0.917044	0.146305	0.217195	0.045249

Table 3. Mean of Specific Crime per Student

Table 3 above offers further analysis into the five specific crimes that happened to students. For crimes like break-in, being stolen, vehicle theft, and crime with weapon support our initial findings that on-campus students experience lower average crime per student. Especially for 'Mean Break-ins' and 'Mean Crimes with Weapon', on-campus students has significantly lower mean than those for off-campus students.

Some aspects that might contribute to this finding is that the broader exposure that comes with living off-campus and the diversity of off-campus housing environments. And, for on-campus students, the more controlled and security-enhanced environment of on-campus housing may contribute to this reduced rate. This leads to our next analysis about types of housing for both off-campus and on-campus students.

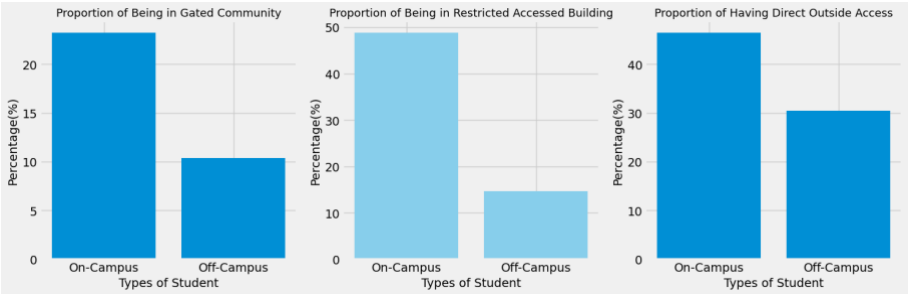

Figure 3. Percentage of Students in Different Type of Housing

Figure 3 above presents percentage of students living in certain types of housing. It is evident that bigger percentage of on-campus students live in gated community (23%) and are in restricted accessed building (48%). This is true since a big portion of college students live in a college dorms that usually have high security system. On contrary, only around 10% of off-campus students live in gated community and 14% of them living in restricted accessed building. The significance difference in proportion of their housing type might explain the difference average crime per student in the Table 2 above. Gated communities and restricted access buildings are often associated with enhanced security measures, which could prevent from criminal activities and consequently result in a lower average crime rate among on-campus students.

However, the third bar chart (Percentage of Having Direct Outside Access) seems to contradict our initial findings since intuitively having direct outside access means being more exposed with crime activities. We expected to see on-campus students to have smaller percentage in this third bar chart, but it turns out that 45% of on-campus students have a direct outside access which is higher than off-campus student percentage which only lies on 30%. Therefore, types of housing is not the only factor that determine the average crime per student.

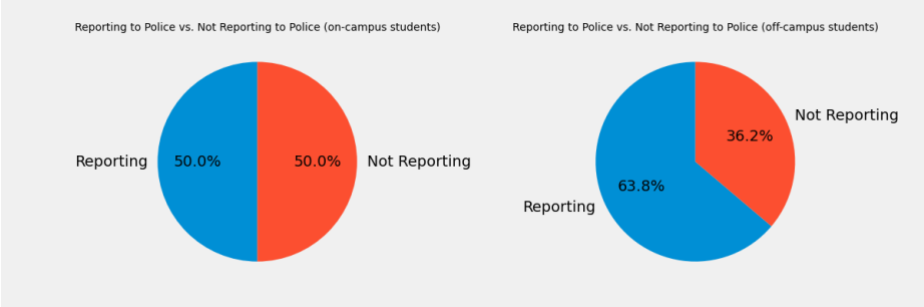

Figure 4. Police Reporting Tendency

Figure 4 above represents pie charts of the times students report and not report incidents that happened to them which they thought was a crime. There is a substantial difference in reporting behaviours of on and off campus students. The statistics reveals that on-campus students has lower tendency to report an incident that happened to them (50%) than off-campus students (63.8%). The relatively secure environment and the presence of campus security measures may deter criminal activities, resulting in a lower incidence of crimes. Consequently, when incidents do occur, on-campus students might be less likely to recognize them as crimes, leading to underreporting. In contrast, off-campus students, living in a potentially less restricted and gated environment, might be more vigilant and proactive in reporting to ensure their safety. Of course, there are several variables unknown that can actually add on to this finding which will be slightly covered in the next section.

	Student	Proportion Offender Known (%)
0	On Campus	73.684211
1	Off Campus	8.037578

Table 4. Proportion of Known Offender

The Table 4 above shows that on-campus students have a significantly higher proportion of knowing the offender. In other words, on-campus students have higher chance to be victims of a crime by someone they know. This correlates with the living arrangements bar charts above, where the reason for the offender being more likely someone they know is because many on-campus students reside in a restricted and gated community. While the environment does reduce the overall probability of falling victim to crime, it simultaneously increases the likelihood of being victimized by someone in the same area/building, higher chance being someone they know.

D. Context

After analysing several college-related variables and their correlation with crime-related variables, we found causality relationships between some of them. The types of housing that the students choose would have an effect with their exposure to crime. It is shown in Figure 3 where on-campus students have a higher percentage of being in a gated community and restricted accessed building (such as a college dorm) which we can assume that these types of housing have a higher security. The findings in the Figure 3 thus has a potential causality relationship with the finding in Table 2 and Table 3 as to explain why on-campus student constantly have lower average of crimes per student than that of off-campus students. Another causality relationship exists between the types of housing variables and the number of known offenders variable. We observe a positive trend, wherein the higher the proportion of students living in gated and restricted communities, the greater the chance of being a victim of a crime committed by someone they know.

Another prominent finding in this report is that on-campus student is less likely to report a crime to the police than off-campus students. There is a possible association between the mean crime per student with this reporting behaviour where the lower the mean of crime per student is the less likely they will report a crime to a police. However, there are arguably possible confounding variables in this finding such as crime severity and awareness of reporting procedures. If on-campus students are exposed to less severe crimes or if certain types of crimes are more prevalent off-campus, this could impact reporting behaviour. Similarly, if there is difference in awareness in the procedures of reporting to police between on and off campus students, it can be a lurking variable that connect the dots of being an on/off campus student with their behaviour in reporting crime. Therefore, to further deep into this question, we need some study or data about perception on reporting and also a variable about crime severity that they had. These variables will provide a more comprehensive understanding of the observed differences in reporting habits between on and off-campus students.

Conditioning on on-campus students variables, we see a positive correlation between being in gated community and having direct outside access (Figure 3). However, intuitively, there should be a negative correlation between those two where if you are in a gated community that you are less likely to have a direct outside access. This can possible a type of collider variable as when we condition on on-campus student, there is this spurious correlation between GATED and OUTSIDE variables.

E. Conclusion

Working on the NCVS 2020 dataset, I learned that types of housing might affect students' exposure to crime activities shown by the difference in average crime per students between on-campus and off-campus students (Table 2). I also learned about intricate correlation between housing types, crime exposure, and students' tendency to report a crime. There are several other variables needed for this study to be more accurate. As a researcher, I hope that this report will contribute some insight to the ongoing discourse on students' victimization.