

# 6.1 Sourcing data

## Data Source

- Each Police force in the UK record the reported crimes. General information on street crime for the past 3 years is publicly available from <https://data.police.uk/data/> , where it can be downloaded split into files for each constabulary by month. An already merged version of this data was available here: [https://www.kaggle.com/datasets/tantable/all-uk-police-street-crime-102018-to-092021?resource=download&select=UK\\_Police\\_Street\\_Crime\\_2018-10-01\\_to\\_2021\\_09\\_31.csv](https://www.kaggle.com/datasets/tantable/all-uk-police-street-crime-102018-to-092021?resource=download&select=UK_Police_Street_Crime_2018-10-01_to_2021_09_31.csv) as well as data on Police officer staffing levels from 2021, which was appended with officer numbers from previous years to match the crime data timeline accessed here: <https://www.gov.uk/government/collections/police-workforce-england-and-wales> As both datasets are from official public government sources, they can be relied upon to be correct.
- This data is fairly recent covering the last 3 years and fairly large with nearly 19m observations. It also fits the requirements given for this project and I am interested to explore street crime in the UK before and during the pandemic, to see if it had any impact, especially during the lockdowns.

## Data profile

- There were no duplicates the in the datasets.
- Several variables in both datasets had missing values. Most notably the Outcome category with nearly 5m missing entries. As this denotes nearly 25% of all observations it should not be dropped and with this being a categorical variable no values were imputed.
- After merging the crime and officer data some further missing values emerged, especially in the crime aggregate and ratio variables. Which were replaced with monthly aggregates, due to the non-equal time frames covered in any one year.
- **Consider limitations and ethics.** Outline any limitations and ethical considerations presented by the content of your data, its source, and/or how it was collected.  
The data is cleared of any PII and no involved person's could be identified by the

observations. As this is made available from each force's database, collection error would be biased through the reports of crimes made by the public or what crimes officers reacted to when on patrol.

- The data is limited in the types of crimes reported in this dataset, as it only looks at street crime.

## Questions

1. What is the relationship between the size of the population and the number of officers in a constabulary?
2. Was there a significant effect on the number of crimes reported during the lockdowns compared to pre-pandemic times?
3. Are certain types of crime more prevalent in specific regions?
4. What types of crime are most likely to lead to the offender being sentenced?
5. What is the relationship between the number of officers and the number of crimes?