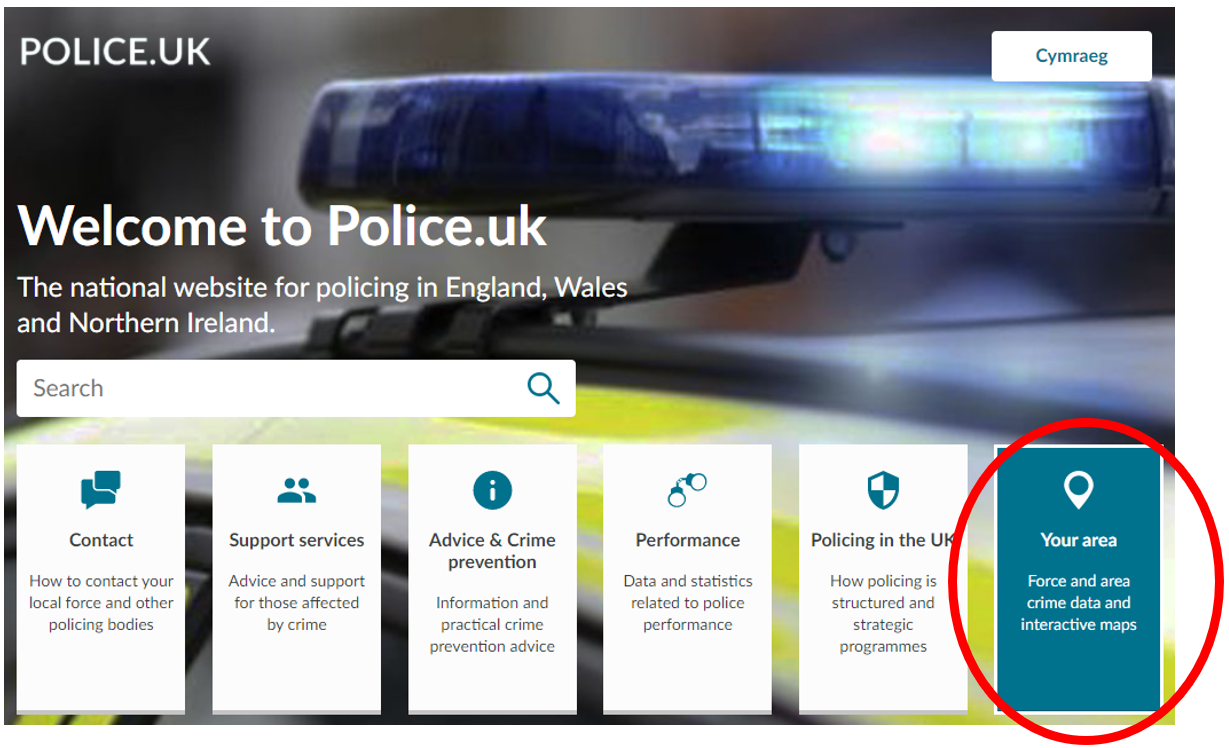
**Tutorial: SOUTH WALES POLICE CRIME DATA**

The aim of this exercise is:

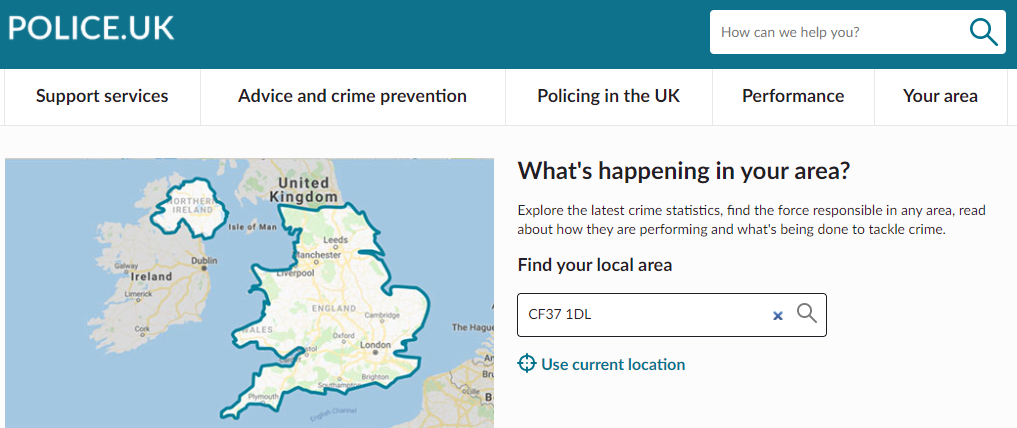
1. ***Familiarise yourself with the Recorded Crime Data for Wales & England.***
2. ***Get you to Download actual Crime Data from the data.police.uk website;***
3. ***Develop some IT skills for managing and manipulating these datasets for any given time period;***
4. ***Load into Excel and Access to query and visualise the crime data (so you can get an idea of how this data may be used – and what you could try to do inside Python).***

This can form the basis of your Python Coursework. The overall goal is to be able to summarise the raw crime data as useful and relevant information.

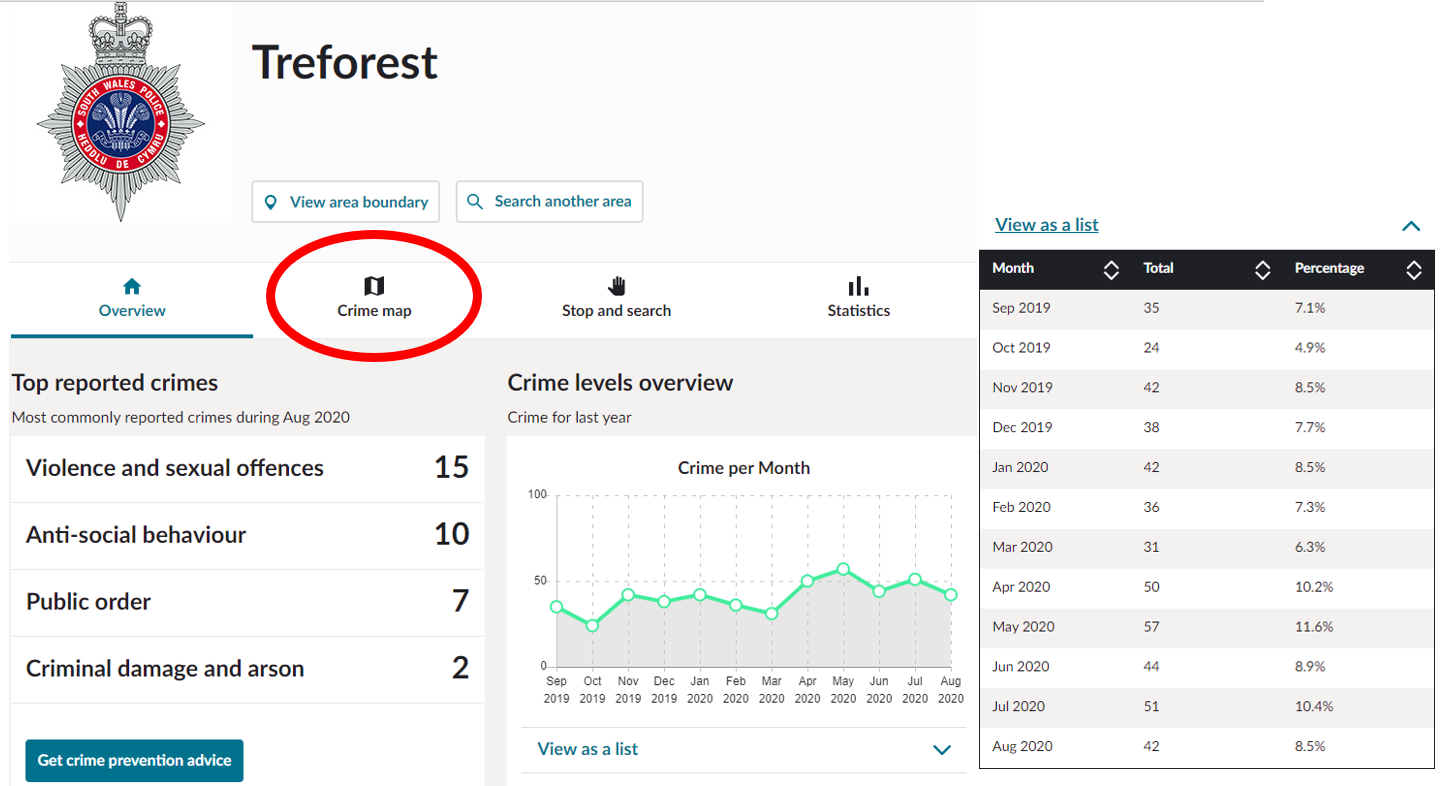
**STEP 1: Open a Web Browser and go to ...** [**https://www.police.uk/**](https://www.police.uk/)



Click the **YOUR AREA** button and type in your postcode, e.g. CF37 1DL (University Postcode) …

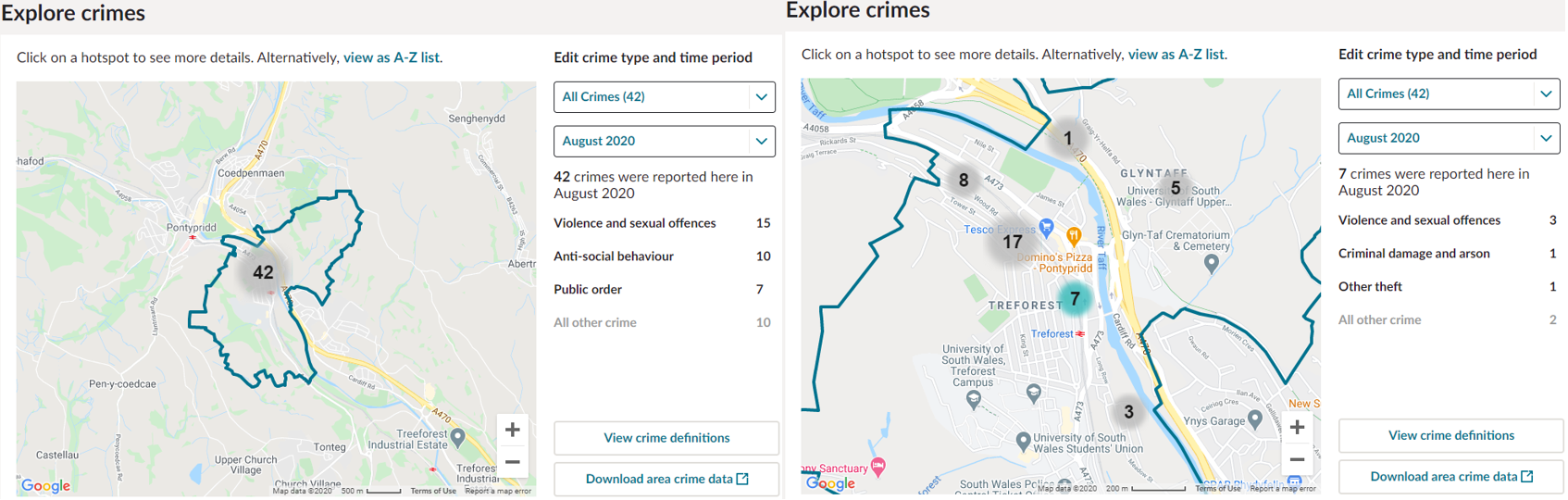


The default Overview tab presents the summary data for the last full month and a crime level overview for the year. Explore your area …



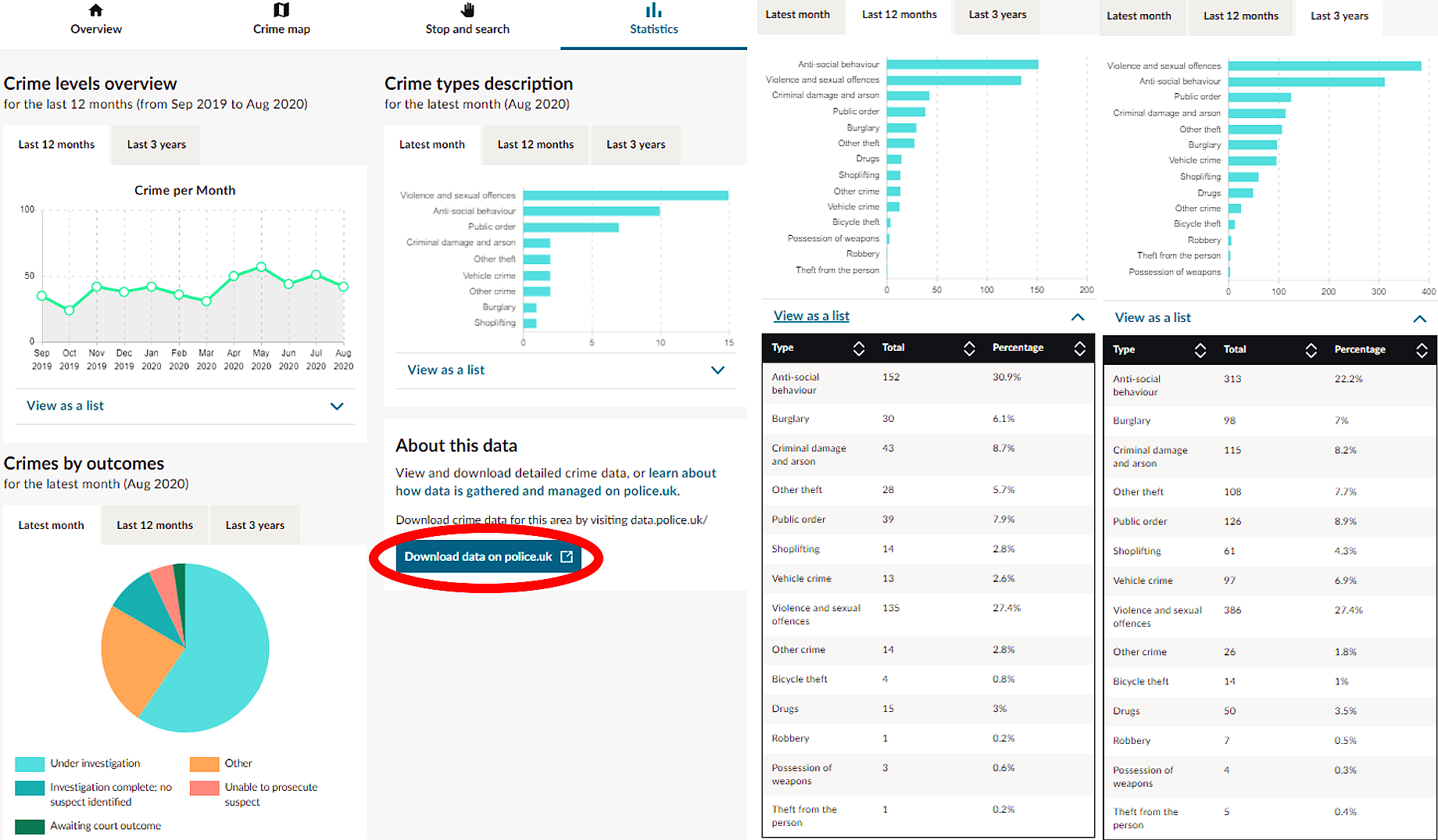
Now, **Explore the crime map** and Zoom in and out of your neighbourhood. Click the hot spot numbers to get a breakdown of the crime types in your area.

Explore the different types of crime and how this compares with the same period last year or the year before. Is overall crime increasing or decreasing? Is serious crime increasing or decreasing? Where are the crime “hotspots” in your neighbourhood.? (Change the dropdown boxes for different months or for specific crime types).



This is a very useful way of examining crime in your neighbourhood. Can you think of a better way of representing this information (e.g. an alternative to manually selecting each month? A time-series animation?).

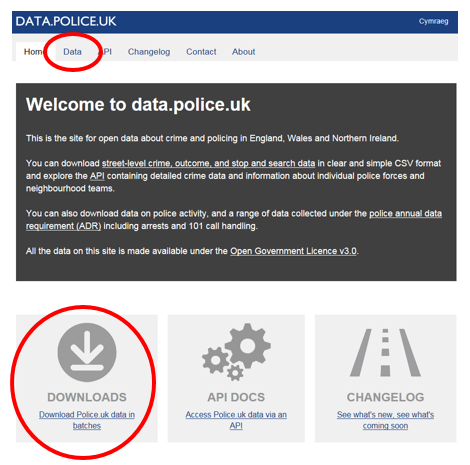
Once you’ve finished exploring the crime map data, click the **Statistics** Tab …



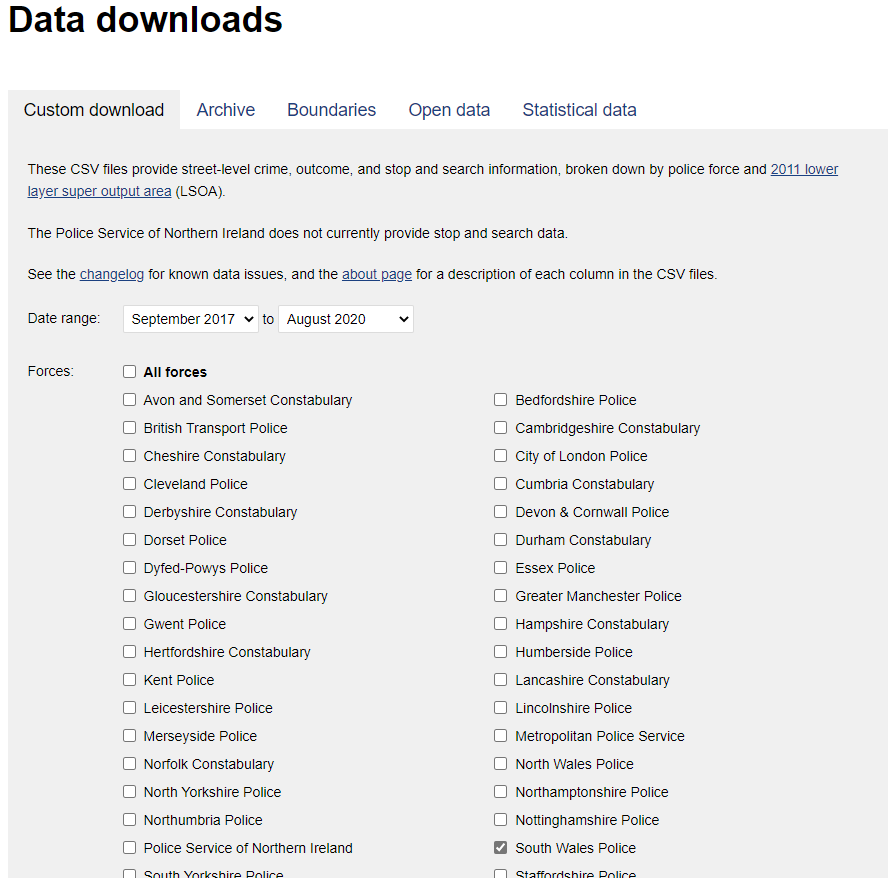
And explore the summary statistics by month, by year and by 3 years (including a break down of all crime types by quantity). Which crimes have increased in the last month or year? Also explore the Crimes by Outcomes and the Pie Charts. What information is useful? What other ways can this information be presented to a reader, a data analyst, a member of the public?

Explore another area (e.g. home postcode) and determine whether the number of crimes for this second area are similar. Is there a better way of comparing two different areas? What is wrong with comparing two different areas by **Total** Number of Crimes?

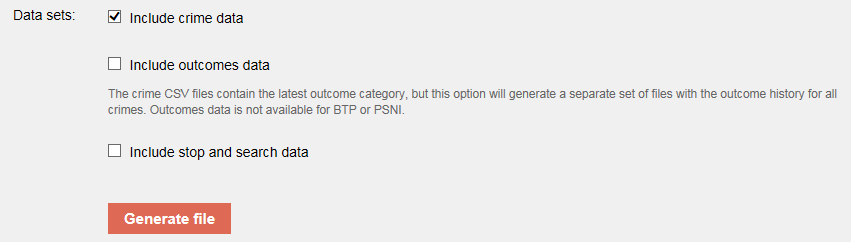
**STEP 2:** Go to the main data download page of the site (either click the link for **Download data on police.uk** or go directly to [**https://data.police.uk/**](https://data.police.uk/) and then either click the **DOWNLOADS** link or the **Data** tab …



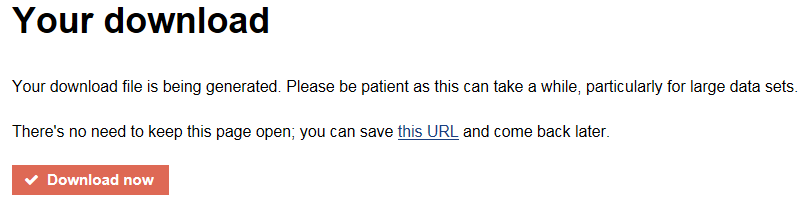
And you will notice that you can download Crime Data (month-by-month) by any Police Authority (or All Forces) at Street level for a chosen Date Range (which gets updated monthly). Select **SOUTH WALES POLICE** and set your choice of Date Range: (e.g. **SEPTEMBER 2017** to **AUGUST 2020**) as shown below:



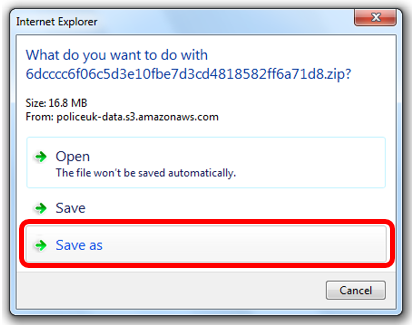
Tick the box for **Include Crime Data** (don’t bother with the Outcomes or Stop & Search data for now) and then **GENERATE FILE**.



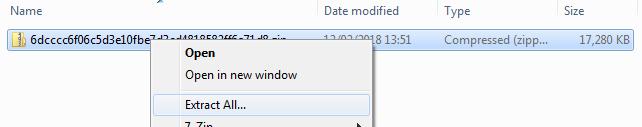
And then click **DOWNLOAD NOW:**



And then click **SAVE AS:**

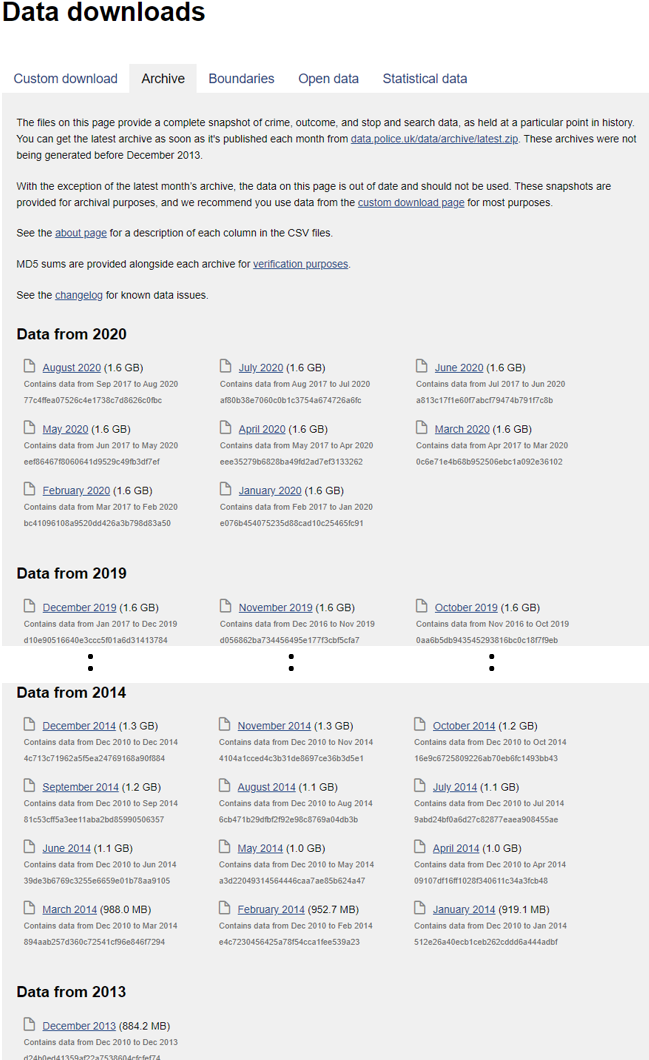


And save the ZIP file to your workspace / USB Stick / Hard Drive / web space etc. Otherwise, save to C:\users\your\_enrolment\_number. In the folder you have saved the downloaded data, right-click on the ZIP file and **EXTRACT ALL** to the same workspace in a folder (and call it something simple, like DATA).



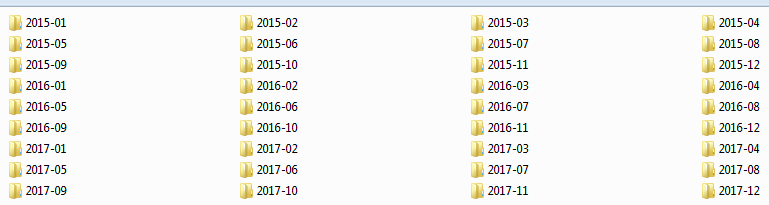
Please note that whilst the last 3 years of data is available to you via the Custom Downloads, you can still access the historical data beyond this time frame by clicking the **Archive Tab …**

Some of these data archives are extremely large (as they cover the whole of England, Wales & NI). You may find some of this data useful if you wish to explore more historical data for your analysis.



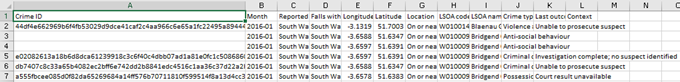
**STEP 3: Explore the data:**

Our unzipped folder will now contain 36 sub-directories corresponding to the 36 months of data downloaded. Each sub-directory will include the CSV (Comma Separated Value) file of the crime data for that month … (N.B. Your data folders will be 2017-09 to 2020-08).





For the time being, open one of these up in Microsoft Excel and investigate the data.



How many crime records are there in your file?

11546

Can you identify all the different types of crime which are recorded?

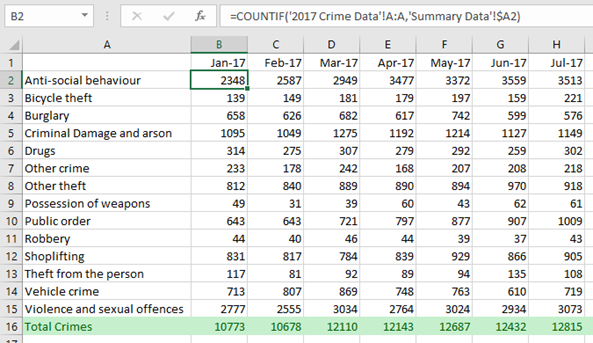
|  |
| --- |
| Anti-social behaviour |
| Bicycle Theft |
| Burglary |
| Criminal damage and arson |
| Drugs |
| Other crime |
| Other theft |
| Possession of weapons |
| Public Order |
| Robbery |
| Shoplifting |
| Theft from the person |
| Vehicle crime |
| Violence and sexual offences |

What crime type is the most common? -

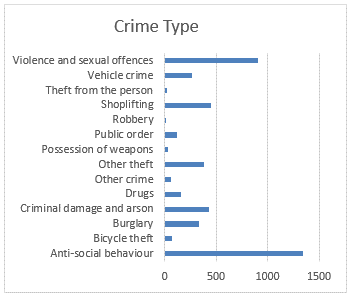
Violence and sexual offences

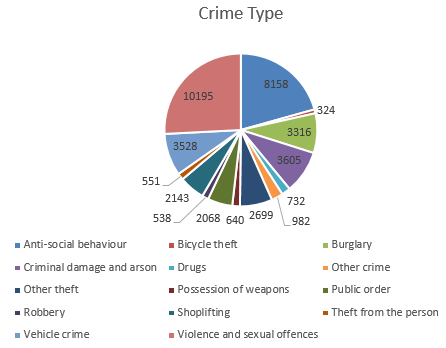


It might be worthwhile seeing if you can quickly summarise the count of these crime types for a specific month (take a look at the COUNTIF function):

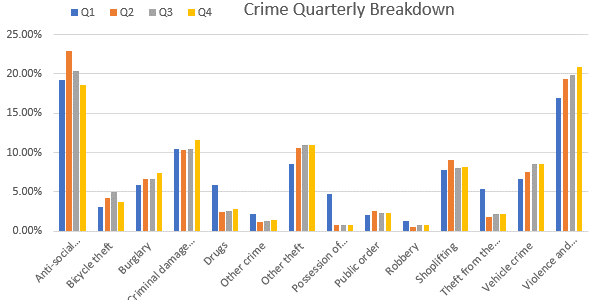


Have a think about whether it would be possible to create similar sort of output in Python? In Excel, also take a look at some of the Chart plotting functions. For example, can you visualise some of this data as:

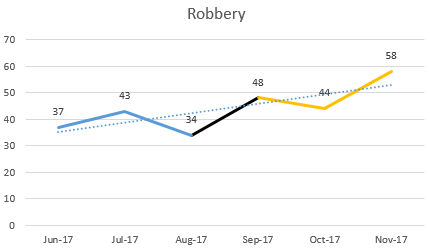
 or

 or

Or

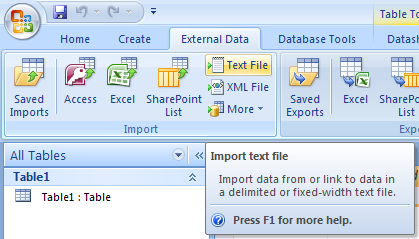


or

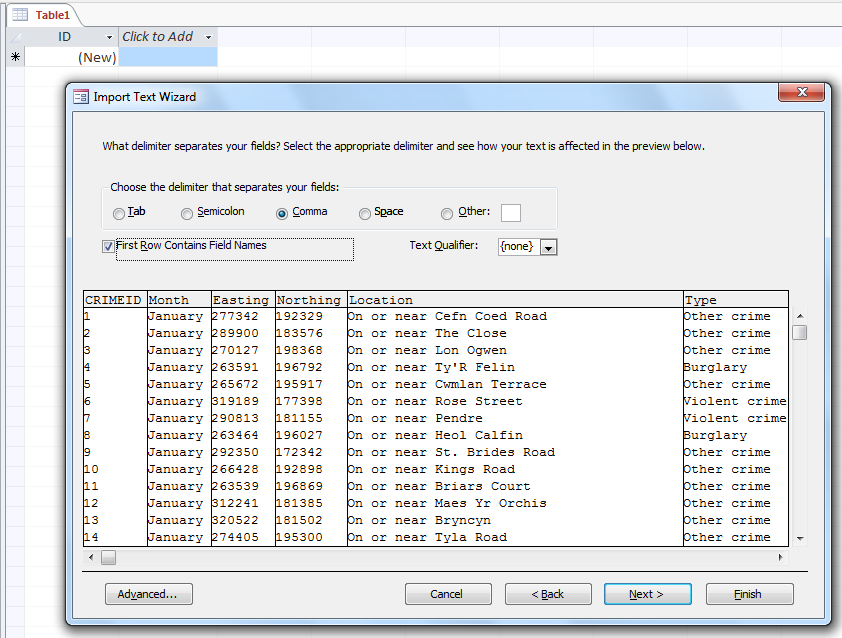


You might hopefully get some ideas for the different approaches for visualising the data (which we will be exploring in the MATPLOTLIB module of Python next time). Also, think about different hypotheses which you might be able to test in Python, for example, “is Robbery on the increase”, or “is Burglary more prevalent in the Summer or Winter months?”, or “What Crimes increased or decreased during Lockdown?”.

If you are more comfortable with using a database such as Microsoft Access, then load the data directly in Access and explore with some SQL type queries …

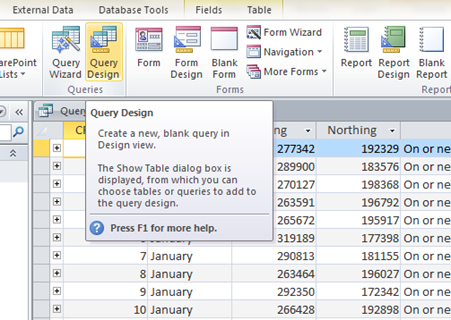


Ensure that in the Import Text Wizard, your delimiter is set to a ***Comma***, and that you check the box for “***First Row Contains Field Names***” (see below) ... and Click ***NEXT*** ...

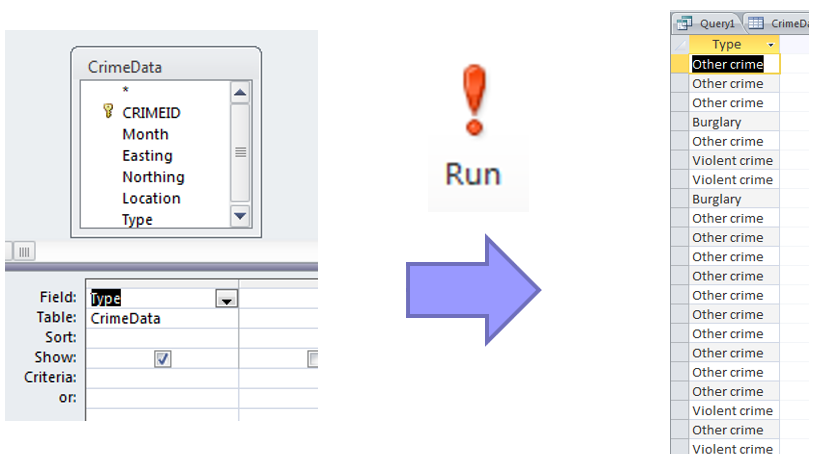


And Click **Next** again. Choose your own primary key (CrimeID)...and generate the Table for this dataset.

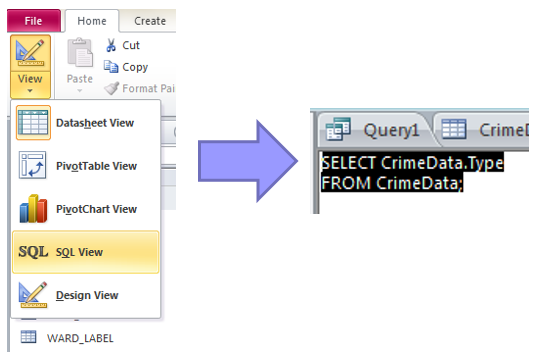
So, in Access, let’s start our query by using the **Query Design** tool …



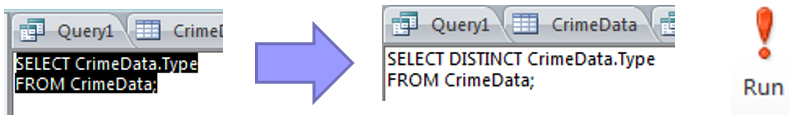
Add your **CrimeData** table, and let’s specify **CrimeData.Type** as the field to be displayed. Run the query



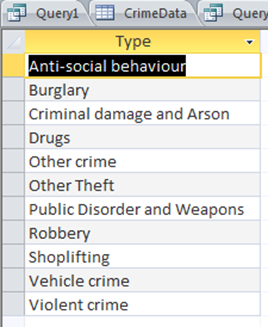
The query actually returns the Type of Crime for all of our 157,980 records (or whatever number you have for Sep2017-Aug2020). However, we just need to identify the unique or **DISTINCT** crime types. So, we’ll need to amend our query. Let’s take a look at our existing query in SQL …



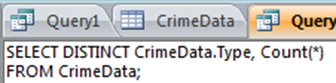
Amend the query to …



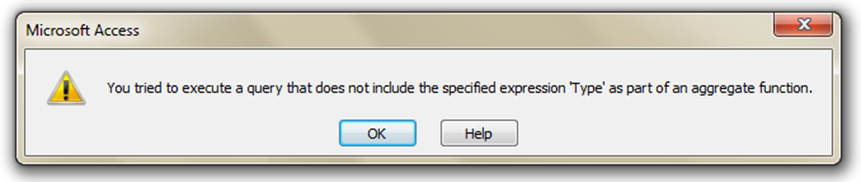
And Run it again. This time your output should look like …



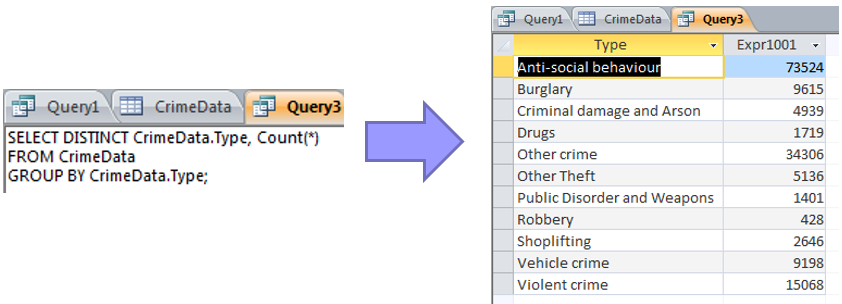
i.e. just the DISTINCT types of crime. We have the first column of our required output table. We now need to find or COUNT the number of occurrences of each DISTINCT crime type. From our notes above, we know that the COUNT(\*) function gives us the number of rows (or records) satisfying a condition. The temptation is to add this to our existing query …



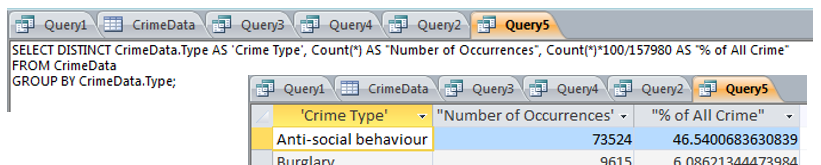
But when we run this, Access gives us the very helpful message …



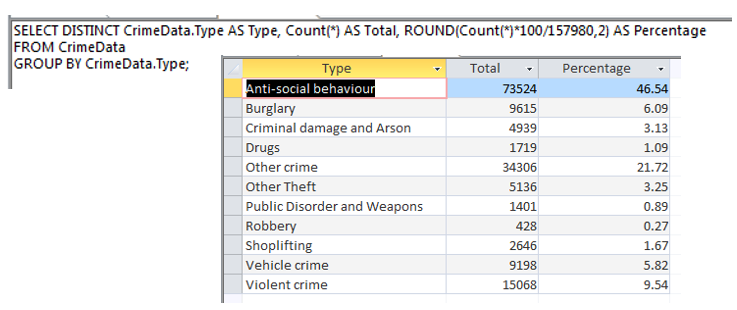
IN SQL, the **GROUP BY** statement is often used with aggregate functions (COUNT, MAX, MIN, SUM, AVG) to group the result-set by one or more columns, which in our above example is the CrimeData.Type. Amend your query (as shown below) – and REMEMBER to ADJUST THE “;” accordingly.



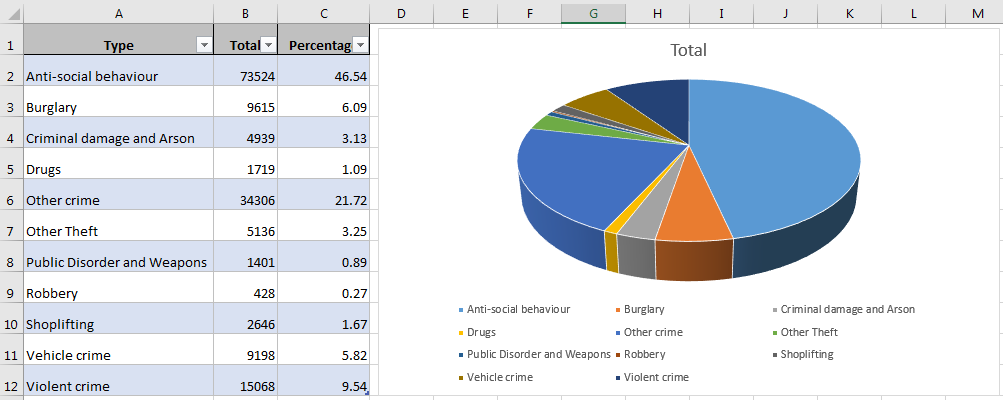
Now amend the query to add the percentage of all crime for which this count contributes … You may have to extend the width of the output column to see all of this data (rather than #######). Now, the column names Expr1001 and Expr1002 aren’t very informative or helpful, so we could rename them (as part of the query) …



If the renamed Column is a single word, e.g. %, then we don’t need single or double quotes around it. Also, the number of decimal places is a bit silly, so we should round down to 2 or 3 decimal places. In SQL, we can do this with the ROUND command …



You can then copy these results into other software such as Microsoft Word or Excel …



How does this chart compare with another police force area for the same month?

How does this chart compare with 12 months earlier?

**BOTTOM LINE …. Explore the data and get some ideas of what it’s all about and what you could possibly do with the data … or what hypotheses you might want to test.**

**You might see some worthwhile patterns or trends in your data which you want to explore further with Python (rather than develop the Python code first – and then realise it’s not worthwhile).**