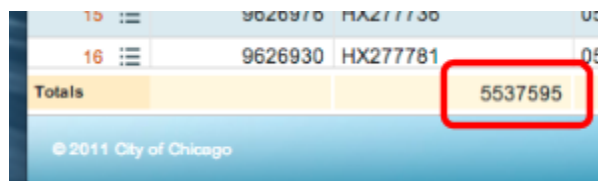


# Visualising “big” data quickly

In this exercise we are going to visualise some data in the Socrata platform. Socrata is an open data publishing platform that allows anyone to interact, filter and visualise data. Significantly this means that large datasets can be manipulated directly within the platform without having to be downloaded. This way you can take advantage of the processing power of the platform and not have to create your own personal data processing environment.

To demonstrate this principle, this exercise looks at the City of Chicago crime data, which is over 5.5m rows of data updated daily (nice!).

<https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2>



|        |         |          |         |
|--------|---------|----------|---------|
| 15     | 9626916 | HA277736 | 05      |
| 16     | 9626930 | HX277781 | 05      |
| Totals |         |          | 5537595 |

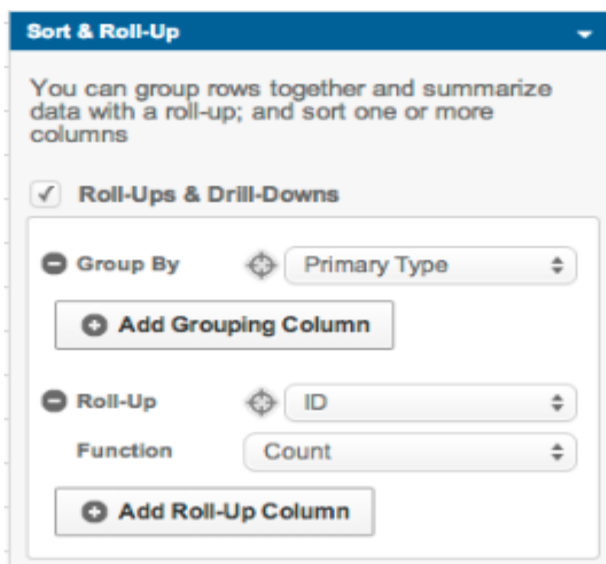
© 2011 City of Chicago

You can see the total number of rows at the bottom of the dataset page.

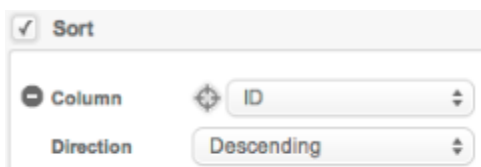
In order to visualise this data it is absolutely necessary to create a pivot table to group the data in some way. In this case we are simply going to group it by the **primary type** column. To create a pivot table first click the **filter** button.



group it  
table first



In the filter panel, the pivot table option is called **sort and roll up** and from here we can simply select the grouping column and the summation column and shown on the left. You might also want to apply a sorting function to order the results with most popular type of crime first (shown below).



Once done click apply and see how long Socrata takes to process 5.5m rows of data.

In order to visualise this new table, click the visualise button (next to the previous Filter button), pick your chart type, e.g. pie chart, select the label (Primary Type) and value (ID) data and socrata will draw you a pie chart. You can then chose to save both this and the table to share with others within the platform. It is also possible to export the data in many formats for processing in other platforms.