The strand unit of Representing and Interpreting Data (Primary School Mathematics Curriculum, 1999) sets out objectives which emphasise the need for children to experience genuine data investigations. Investigations may take the form of a project/problem to be addressed over a number of lessons. The investigation presented here (and used in Limerick City classes) uses the PPDAC cycle (see InTouch, March 2012) as an organising framework. The investigation requires pupils to explore the characteristics of a distribution (set) of data collected as part of an investigation.

Context: Crime Scene Investigators (C.S.I.)

Step 1: Problem (PPDAC)
Introduce the context by telling children ‘Today we’re going to investigate a crime scene’. The CSI theme tune may be played. Explain that an incident has happened in the school and that their class has been selected to help the Gardaí with their investigations. Present an image of the crime scene (e.g. school garden) (image 1a). Provide children with information regarding the crime (e.g. vandalism). Highlight that the only evidence found by the Gardaí crime investigation bureau was a shoe print (image 1b).

Step 2: Plan (PPDAC)
Invite suggestions regarding how to rule themselves out as suspects. Responses may include: ‘place shoes in actual shoe print’; ‘measure shoe sizes’. Explain that statements alone are not conclusive scientific evidence and that the crime scene cannot be tampered with. Tell the children that each pupil is going to find the length of their shoe print as this is the procedure recommended by the Gardaí.

Step 3: Data (PPDAC)
Demonstrate tracing a shoe print on an A4 page (image 2a) and subsequently measuring the length of this shoe print (image 2b). Remind children of the importance of using the ruler/metre strip correctly (i.e. start at 0) as well as the appropriate unit of measure (cm).

Step 4: Analysis (PPDAC)
Following discussion of group predictions, tell children that you are going to investigate the characteristics of the classes’ actual shoe size (i.e. the distribution). Introduce the line plot as a useful graph to represent data. Have the children identify the range of
Using Novel Contexts to Explore Data Sets

Children in our classrooms mentioned: *25cm is the most popular shoe length* *The graph goes from 22 to 27* *It looks like the Eiffel Tower*

Children’s observations should be allowed to guide the formal analysis of the distribution i.e. the arrangement of data values on the line plot. For example, children’s references to the ‘bump’ facilitates introduction of the statistical term for the ‘most popular/common’ as the mode i.e. 25cm.

Inform the class that the range (the interval from the lowest to the highest data value) of the data set is 22-27cm (and not 18-31cm as some children mistakenly think).

It will be necessary to explicitly address many features of the graph using both formal and informal language. Start by discussing the shape of the data. When considering the shape, refer to the location of the cluster of data (naturally occurring group of values from 22-27cm) as well as the fact that there are no outliers (unusual data value separated from the cluster) or gaps (holes) in this data set. Discuss what this means in the context of shoe size.

With older children, then move to identifying measures of central tendency. Children can find the mean shoe length. Remind children of the meaning and formula for the mean (see Intouch, April 2012) and encourage the use of calculators to identify the mean of the data. In this example (image 4) the mean is 25cm. Subsequently the median can be introduced as the exact middle value of the data set (where half of the data are above and half below this value). It is possible to describe the median using the body e.g. ‘With my arms outstretched my head is the median i.e. there would be the same amount of data at either side of it’. Demonstrate the method of finding the median i.e. write the data in order and cross out pairs of data values starting at the ends in order to find the middle value.

In the case of an odd number of values, a single middle value (median) emerges. However, in this example (image 4) because there is an even number of values (22 pupils) there are two middle values of 25cm and 25cm. Therefore the median is 25cm (the midpoint between the two middle values). Explore what the median would be if the two middle value were 24 and 25 (answer: 24 + 25 = 49 / 2 = 24.5cm).

Use the line plot to highlight that the mean, mode and median of the data set are all 25cm because the distribution is almost symmetrical. Allow groups to compare their predictions with these findings. Lead the class to the conclusion that 25cm is the shoe length that best represents the class. If the mean, mode and median are different for your class, allow children to report/vote on the shoe size that best represents the class by considering all these values.

Pose questions that support children in reading beyond the data. Present questions such as ‘What do you think a distribution for 2nd class shoe lengths would look like?’; ‘Does this data represent all 5th classes?’; ‘Who might have a shoe length of 31cm?’ Explore the consequences of an outlier (e.g. ‘X’ over 31cm) being added to the line plot through questions such as ‘Will the mode change?’ (Answer: no); ‘Will the mean change?’; ‘How?’ (Answer: yes; it will be bigger); ‘Will the median change?’; ‘How?’ (Answer: No, it will still be 25cm). It is possible to work through each of these investigations. Ask the pupils what else might change if you measure the shoe sizes of a different sample (Answer: possibly the shape, mean, median, mode).

Step 5: Conclusion (PPDAC)
You may then get a ‘mock’ phone call from the Gardaí telling you that the actual shoe length found in the garden was 28cm (Note: the crime scene shoe length should never be the same as one of the children’s shoe length). Explore what this means in the context of the data collected from your class (Answer: all pupils in this class are innocent). Ask each group to provide suggestions as to the next steps in the investigation. Groups may recommend that the investigation focus on ‘4th and/or 6th class in this school’; ‘1st years in secondary schools’; ‘the school staff’.

The children’s suggestions could be the basis for follow-on work, where data regarding shoe length is examined from another sample e.g. school staff. This activity, in turn, will provide opportunities for children to make comparisons between the respective data sets. The ‘typical’ value will prove invaluable during such an activity, considering that the two data sets to be compared will be different sizes.

Acknowledgements: Sincere thanks to the co-operating pupils and staff in St. Michael’s N.S. (Limerick City) and Scoil Mhathair Dé (Limerick City) as well as to Sáamus Heffernan, Aisling Garrity, Aoife Hill, Freida McGrath, Laura O’Callaghan and Rebecca Larkin for their contribution to Lesson Study.