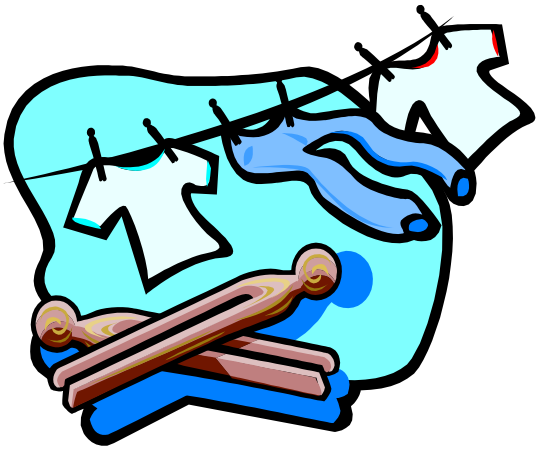
Hanging out the Dirty Data

When you have some data the first thing you need to do is to check it out and get rid of any obviously wrong or false data. This is called "Cleaning the Data".

In the spreadsheet 'Dirty' several playful pupils have been deliberately tampering with the data. See if you can spot which rows and cells have been interfered with and by which of the following characters:

Pointy Pete:	moves Decimal Points around and adds in unnecessary ones.
Obvious Olive:	puts in **very** obvious errors.
Silly Samantha:	thinks it is funny to answer the question. For example: What is your name? - "Donald Duck"
Devious Dave:	thinks up clever ways to change the data

Download the Excel spreadsheet or find a copy of it on page 2.



Can you make some decisions about what to do with the dirty data?

Investigate the Spreadsheet further and try to come up with a

sparkly clean final version!

Note: The data shown overleaf was collected over a number of years, so pupils' in a particular year group may have different birth years.

Hanging out the Dirty Data

Dirty Spreadsheet

Row 1	Boy or Girl	Date of Birth	School Year	Born in	Height	Foot Length	Favourite Subject	Distance to School
Row 2	Boy	12/04/01	5	England	143	26	Art	1-2km
Row 3	Girl	31/02/02	4	England	132	22	Science	less than 2 km
Row 4	Girl	14/01/01	5.00	England	14.2	2.3	PE/Sport	2.5423 km
Row 5	Boy	07/09/05	6	England	136	25	Art	1-2km
Row 6	Boy	13/12/01	4	Wales	128	24	PE/Sport	1-2km
Row 7	Boy	14/03/11	5	England	140	67	PE/Sport	less than 1 km
Row 8	Girl	06/05/01	7	Wales	142	24	Art	3-5km
Row 9	Girl	15/08/04	6	England	138	21	Art	85km
Row 10	Boy	20/02/02	6	England	192	23	PE/Sport	1-2km
Row 11	Girl	19/05/90	6	England	140	20	Maths	1-2km
Row 12	Neither	29/06/03	7	Planet Zog	48	21	Going Home	3000km
Row 13	Boy	09/10/91	4	Scotland	128	21	English	less than 1 km
Row 14	Girl	18/12/02	5	England	135	21	Geography	less than 1 km
Row 15	Girl	18/07/01	0.5	England	13.7	20	Art	3-5km
Row 16	Boy	03/06/34	4	Wales	129	21	Art	less than 1 km
Row 17	Girl	13/02/03	7	Wales	148	23	Art	1-2km
Row 18	Girl	15/09/04	7	England	150	22.5	PE/Sport	1-2km
Row 19	Girl	07/08/03	7	England	140	24	Art	less than 1 km
Row 20	Boy	08/06/03	7	England	142	24	Computing	less than 1 km
Row 21	Boy	31/11/03	11	England	1520	22	Computing	5-10km
Row 22	Both	16/07/04	8	Engand	142	26	Japanese	2-3 km
Row 23	Girl	28/04/04	8	England	145	26.5	PE/Sport	1 mile
Row 24	Boy	25/03/02	4.1	England	132.1	2.4.5	Maths	less than 1 km
Row 25	Boy	26/02/02	4	England	130	21	PE/Sport	less than 1 km
Row 26	Girl	08/07/10	6	Waland	142	22	Art	2-3 km
Row 27	Boy	23/05/01	6	England	151	25.5	Maths	2-3 km
Row 28	Boy	01/03/87	9	England	162	25	PE/Sport	less than 1 km
Row 29	Girl	07/08/01	6	England	150	23	History	2 roads
Row 30	Girl	03/03/92	4	England	135	21	English	less than 1 km