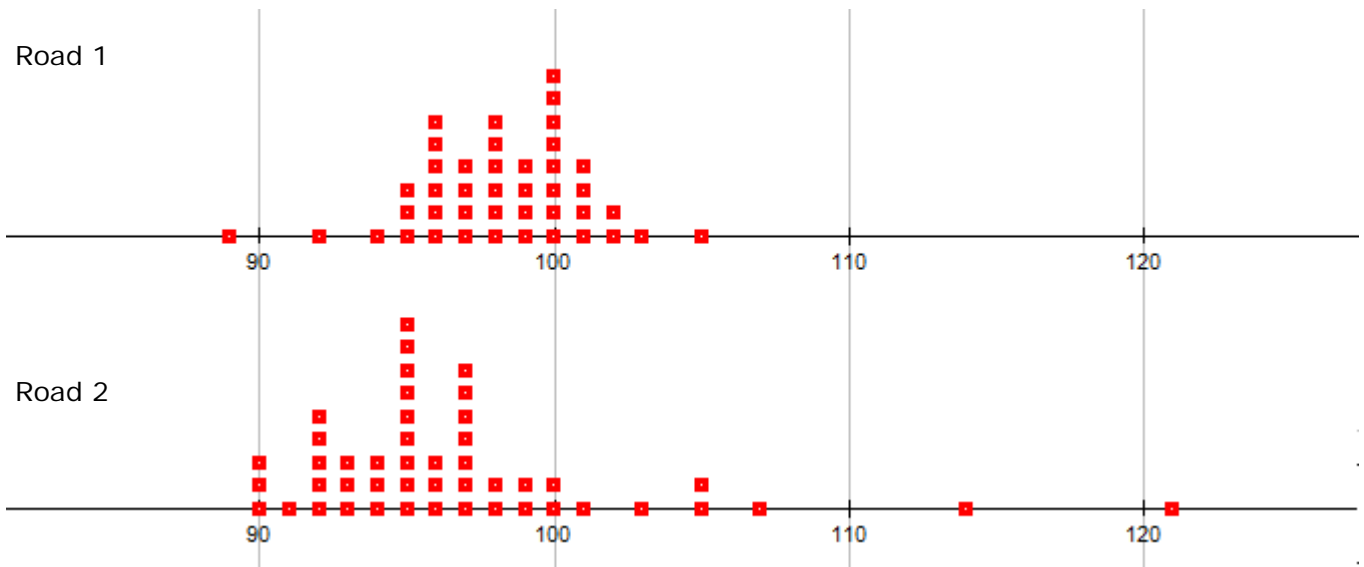# Level 1 Data Practice #4

Betaville Council collects data on the speed of every truck moving along two country roads over one Friday in May. The statistics and a dot plot of the data are shown below.

Statistics Calculated

	Road 1	Road 2
Number of trucks	42	49
Lowest value	89	90
Lower Quartile	96	93
Median	98	95
Upper Quartile	100	98
Highest value	105	121
Mean	98.1	96.9

1. Draw an appropriate graph for the data.

2. Which road has a higher typical speed?

   Give statistical reasons for your answer.

3. How would you describe the distributions of speeds on the two roads.

4. Make three suggestions to improve the quality of the data gathered.

MAN vs MATHS 2014

Answers: Level 1 Data Practice #4

These answers are a guide only, as a range of answers is possible.

1.

Road 1
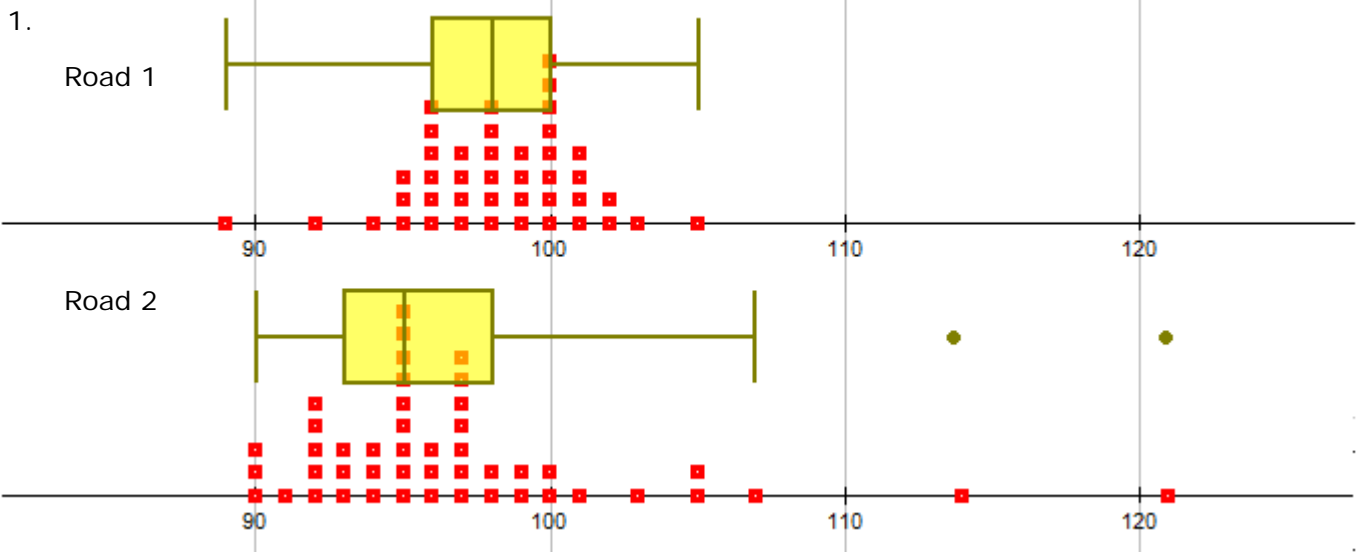
Road 2

The ideal graph is a box-plot, which should look something like the one above.

The whisker of Road 2 can be extended all the way instead, but the dots are better, as those two values are quite a long way from the main group, and to leave them in might give a false impression of how many trucks are going very fast.

2. On the day in question Road 1 has a higher typical speed, as seen by a higher median and a higher mean. I would rely on the medians rather than means, as they are less likely to be affected significantly by extreme values (like the two fast speeds on Road 2).

   The medians are outside the IQRs so it seems unlikely that the difference is due to sample variation. DBM over OVS is $^3/_7$ which is quite significant, much more than $^1/_3$.

   However we cannot make a call about which road has a higher typical speed for any other days, as our sample is not representative enough, being based only on one day

3. Road One has a distribution that is more or less symmetrical about a central peak ("normal"). We see, as we would expect from a small sample, a bit of variation from that.

   Road Two has a distribution that is skewed, with a tail to the right (higher speeds).

   Note: there is no reason to call the two values at 121 and 114 "outliers" as these values are not that uncommon on roads, so may well be normal – we don't have enough data.

   Both roads have similar ranges for typical speeds (IQRS of 4 and 5, and most values within 5 of the medians), although Road Two's main cluster is about 3 km slower.

4. A range of days of the week and times of the year needs to be taken, randomly selected, instead of one day, to ensure a representative sample. Either all day is required, or a random selection of times from each day.

   The average speeds of cars might be compared, to see if it is a problem wider than just trucks.

   Data could be divided into big trucks (say semi-trailers) and small trucks. It is important that both sites were measured using the same definition of "truck".

   (*Note, don't spray multiple suggestions in questions like #4, hoping one will stick. Keep to one more than asked and group in one answer, clearly numbered 1), 2) similar ideas.*)