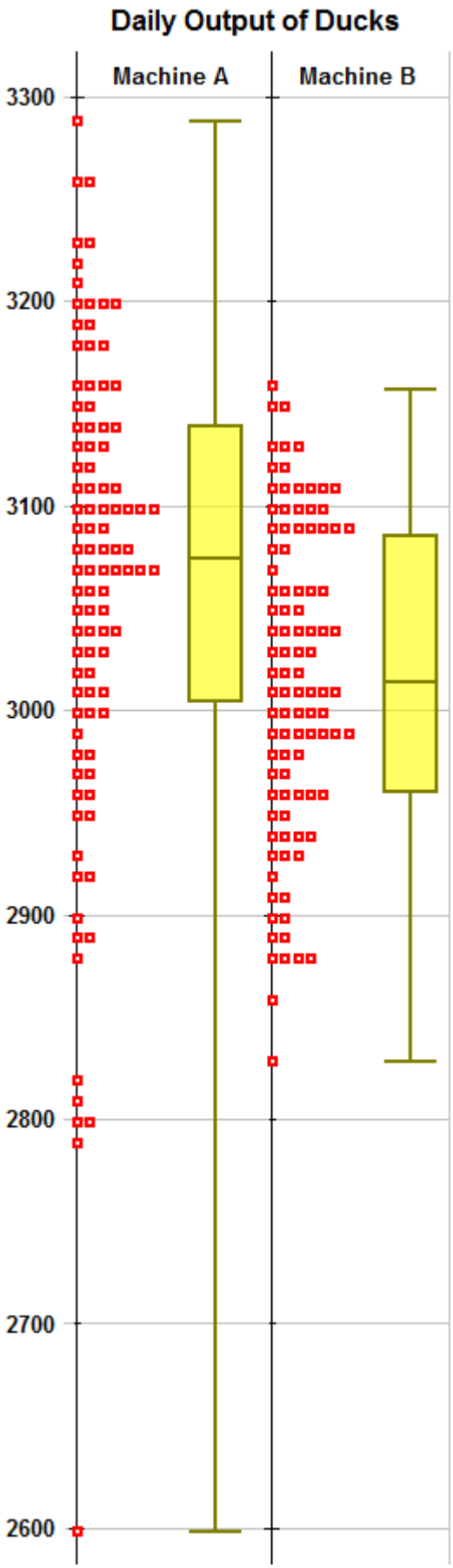
Level 1 Data Practice #10

Alphaville Plastics Company has two machines making plastic ducks. They are thinking of replacing one with a newer model, and need to decide which is the better one to replace.



They count the number of ducks made each day by both machines for 100 working days.

The results are shown in a table below and a combined box plot/dot plot (with the dots rounded to the nearest 10).

Statistics Calculated

	Machine A	Machine B
Lowest	2598	2829
Q1	3005	2962
Median	3075	3014.5
Q3	3139	3085.5
Highest	3288	3157
Mean	3060.9	3016.0

1. Why is the dot plot plotted with rounded data?
2. How might they improve the graph or the calculated statistics?
3. Machine A is more likely to break down. How can you tell this from the data or graph?

   How often do you think it broke down?
4. The company decides that Machine A makes more ducks on average, so it is the better machine. Is their decision justified?

   Give statistical reasons for your answer.

4. If the data was gathered for more days, how might it differ?

   Would doing that help decide which machine was better?

MAN vs MATHS 2014

Answers: Level 1 Data Practice #10

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

1. It is rounded to avoid the dots being placed on top of each so closely that one cannot see the general shape of the distribution. The exact values do not need to be shown, but the general spread does.

2. The result at 2598 for Machine A looks very much like an outlier. It could be shown on the graph with a dot, and the whisker only going to the next point (about 2790).

   If the result at 2598 was a genuine outlier, and unlikely to occur again, it could be excluded from the statistics to avoid getting a false impression of the correct mean. (However in such a large sample size it doesn't make that much difference.)

3. There is a clump of five values at around 2800 that is separated from the main distribution. It makes sense that these are breakdowns, as a breakdown would cause less ducks to be made. The lowest result might well be a major breakdown.

   So say broke down six times (counting the dots for the low clump and lowest value).

   (Even if the lowest values of B are caused by breakdowns, they cannot be as serious as the breakdowns for machine A.)

4. Both the mean and median suggest that Machine A is better.

   The mean of the two distributions is about 45 ducks different. If anything the difference between the two machines is more than this, as the mean for Machine A is dragged down quite a lot by the lowest values – we can see that the median for Machine B is almost identical to its mean, but the median is 15 higher than the mean for Machine A.

   The difference in medians is almost exactly a third of the OVS. For sample sizes of 100 that would suggest that the higher median for Machine A is significant, and the difference is not just due to sample variation (over $^1/_5$ is enough for n = 100).

   (Note: students might say that the medians for the each machine is inside the IQR for the other, so we cannot "make the call". This might be given some marks, but such as statement is at a lower level than expected for Year 11 Excellence.)

   The difference in means is not as much as the difference in medians. Means are more useful in this case (as we care about the overall number of ducks made, not the amount on the typical day) but the difference is still $^1/_4$ of OVS, so that suggests the difference is still significant, and we can make the call that Machine A is better.

5. The sample sizes are already 100, which decreases a lot the effect of sample variation compared to a small sample.

   The most likely result of more data is that the calculated statistics would stay mostly about the same. If that happened it would increase our confidence that Machine A is better (because as sample sizes increase we can make the call based on smaller differences in medians).

   There is a chance however that the extra data would give a smaller difference of means and medians. If that happened it might reduce our confidence that Machine A is better.

MAN vs MATHS 2014