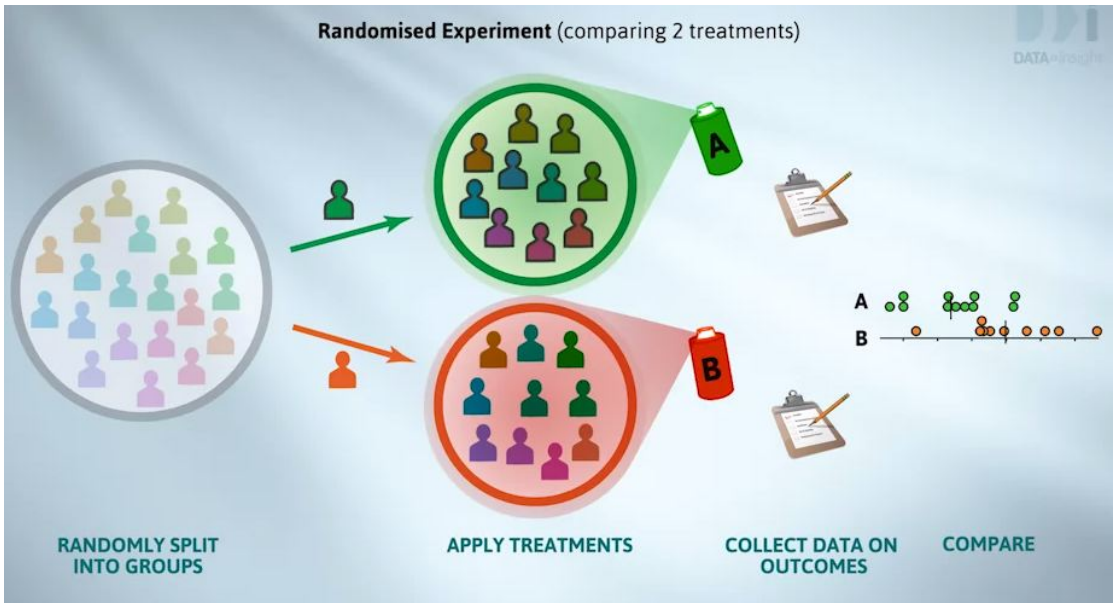
THE UNIVERSITY OF AUCKLAND
NEW ZEALAND
Te Whare Wānanga o Tāmaki Makaurau

WEEK 7

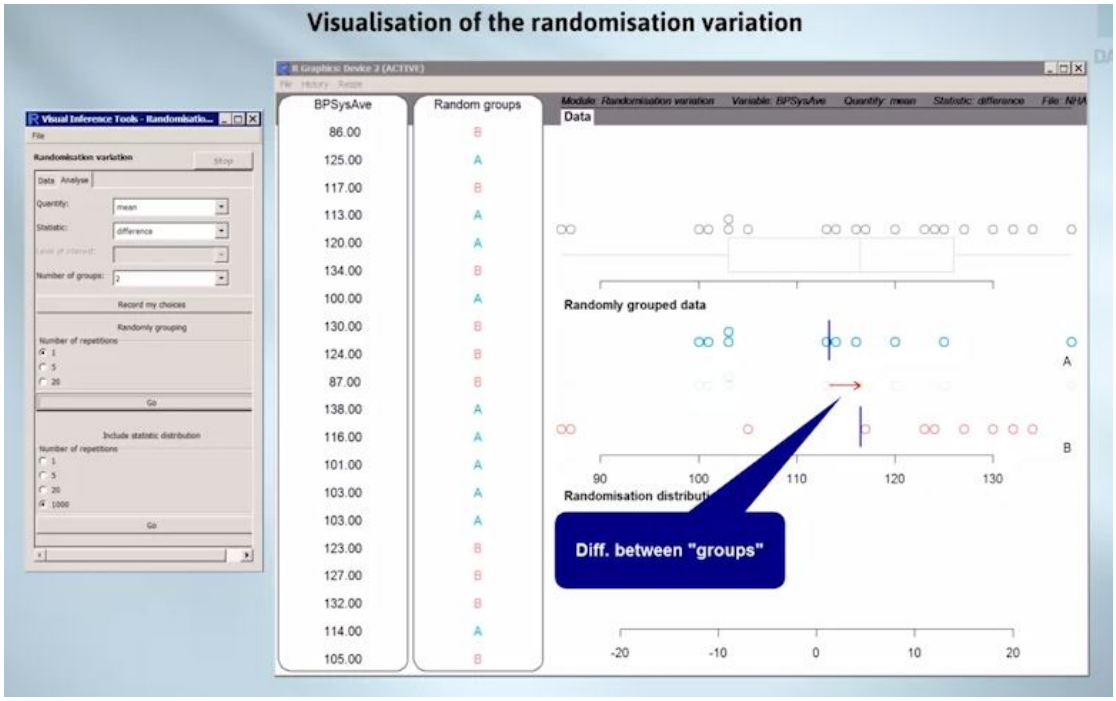
RANDOMISATION VARIATION by Chris Wild

In the last video and the readings, we saw how random assignment of experimental units to treatment groups is the most reliable way people know for constructing fair tests of treatments. I ended the last video by saying that although random assignment is best, it's by no means perfect. I'm now going to show you the main problem with it. And while we're learning about that, we'll sow the seeds of a solution.

Keep this simple scenario in the back of your mind. We want to compare a new drug that's supposed to lower blood pressure to a control treatment.
We have 20 people who we randomised to receive the drug treatment or the control treatment. And some time later, we measure their blood pressures.



If the drug's effective in lowering blood pressures, we'd expect the drug-treated group to have lower blood pressures on average than the control group.



Example still from random variation visualisation animations

Here are the systolic blood pressures of 20 people in the NHANES data set. I'm pulling a copy of the points down into the middle panel. Now I've randomly labelled and coloured half of them as belonging to Group A and half as belonging to Group B. Now I've pulled them apart into two artificial groups, marked the position of the mean blood pressure of each group, and I've shown the differences between these two means with a red arrow.

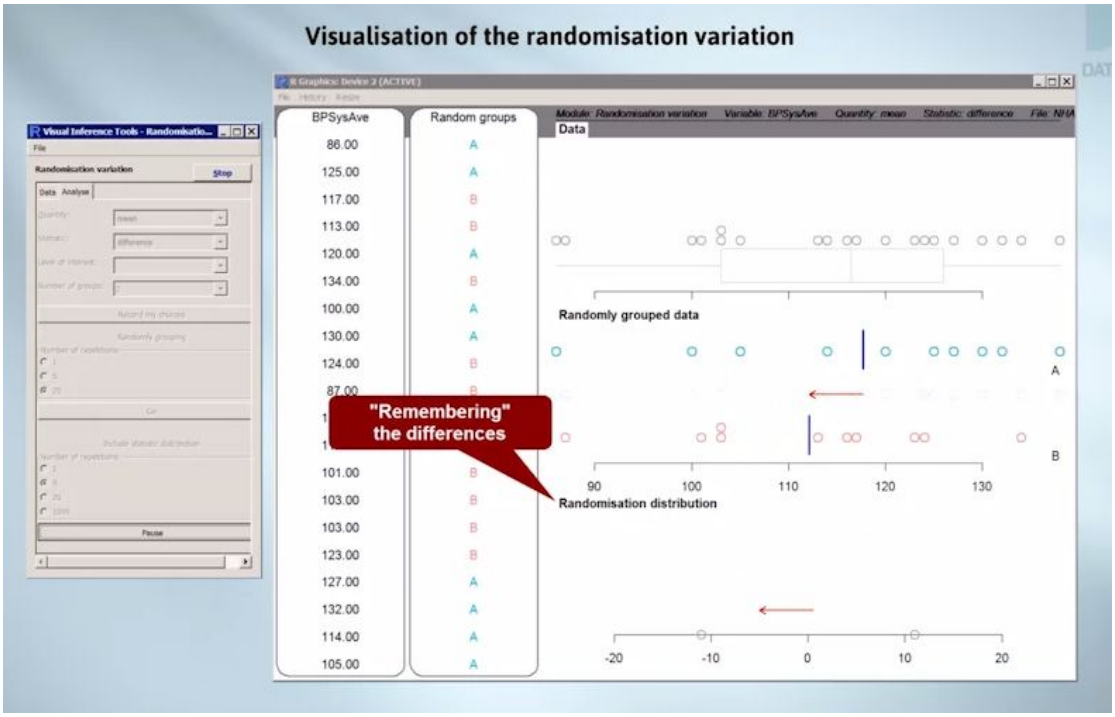
The B Group mean is bigger than the A Group mean by about three units. These groupings are entirely random and artificial. Nothing's been done to anyone. We've just randomly divided the available people into two groups.

Let's do it again-- bring them down, randomly label half of them as belonging to Group A and the other half as belonging to Group B. Separate the groups and show the difference in means. The difference is really small this time. I'll do it faster a few more times to reinforce what we're doing.

Now I'll do it 20 times. Did you notice that sometimes there were quite big differences between my randomly chosen groups? Watch again. Look at this one.

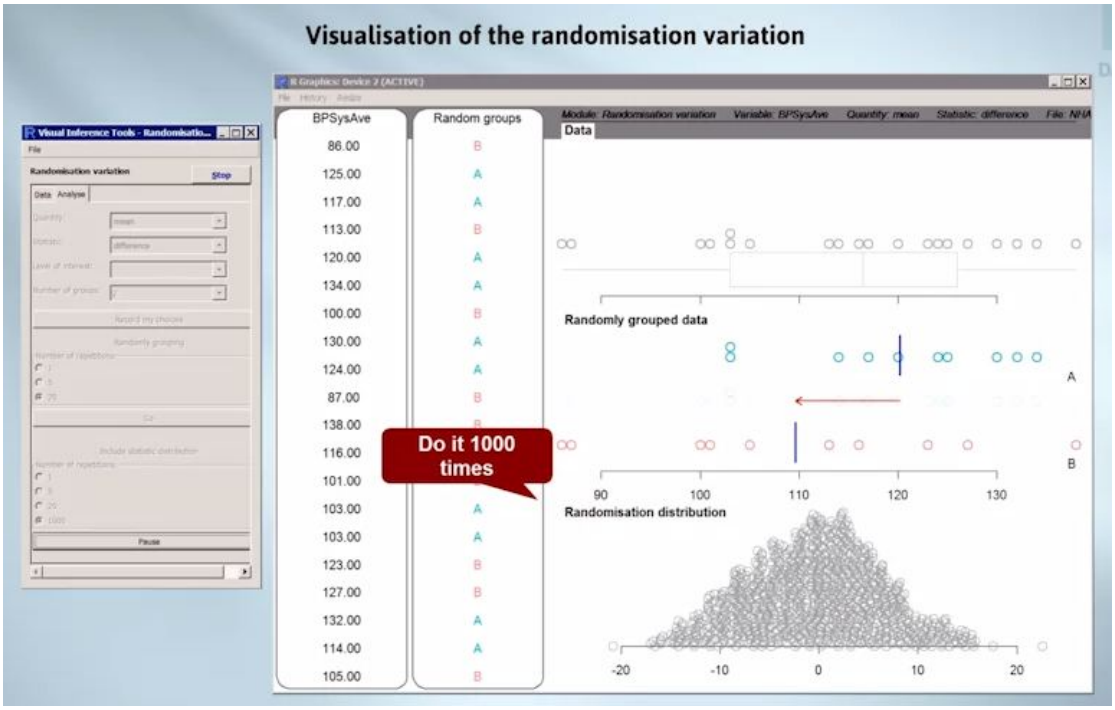
Treatment B is clearly reducing blood pressure. But of course, it's not. Treatment B is not a treatment at all. What we're seeing is just the result of random labelling.

We've been seeing random labelling with A's and B's producing small and large differences in mean blood pressure. But we need a way to record these differences as they go by.



Example still from random variation visualisation animations for remembering differences

This is how we do it. The arrows drop down, and we record the links in the bottom panel. Right pointing arrows (positive differences) go to the right of zero, while left pointing arrows (negative differences) go to the left of zero.



Now let's do this 1,000 times. We build up a distribution of differences. Some of them are quiet big. We're seeing differences in means of up to about 16 units in either direction with the occasional one even larger. This is quite large even compared with the person to person differences in blood pressures.

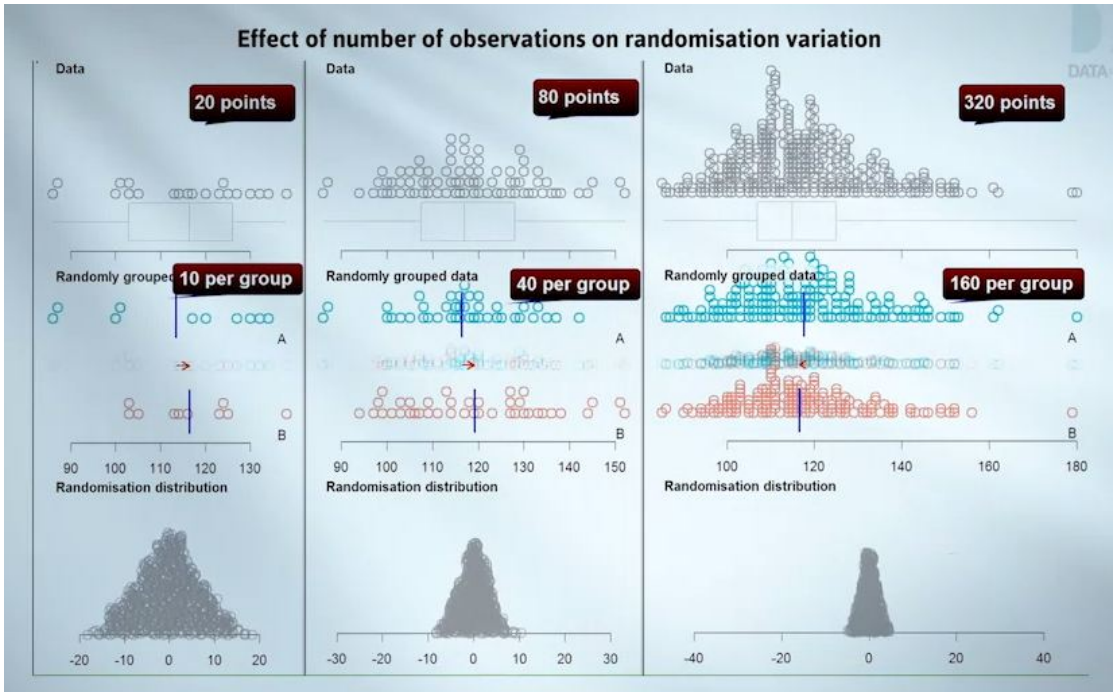
What's the moral of all of this? Random assignment to groups acting on its own can produce quite large differences between group centres. There were no treatment differences here. It was random assignment and nothing else. Let's go back to the statement from the set-up scenario.

If the drug is effective in lowering blood pressures, we'd expect the drug treated group to have lower blood pressures on average than the control group-- so far so good.

We now know, however, that random assignment to groups can, on its own, produce quite large differences between group centres. If we saw a difference between the drug and control groups in experimental data of a size that could easily be generated by random assignment acting alone, this would not provide any evidence that the drug was effective.

So to believe that we have evidence that the drug is effective, we'd have to see differences between the drug group and the control group that were bigger than

random assignment acting alone would produce. In the next video, we'll flesh out this idea and show how we can use it in data analysis.
The type of variation we've been seeing here is called randomisation variation. And we've been using the randomisation variation module in VIT. Just as with sampling error, randomisation variation becomes less of a problem when we have more observations. Here we're comparing randomisation variation with 10 per group and 40 per group. Here I've added 320 points.



As with sampling, every time we multiply the number of points by four, we halve the amount of randomisation variation. But whereas major clinical trials in medicine may have thousands of observations, and thus low randomisation variation, experiments on industrial processes, where obtaining each observation can be hugely expensive, tend to have small numbers of observations.

This brings us to the end of this video. Next time, we'll talk about the randomisation test, a statistical procedure for assessing the plausibility of a "Your results could just be due to chance" explanation.