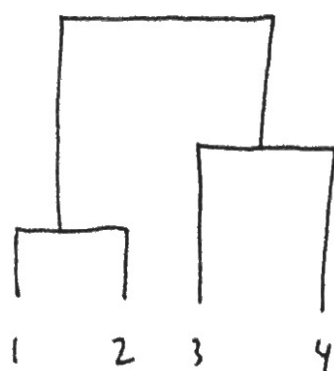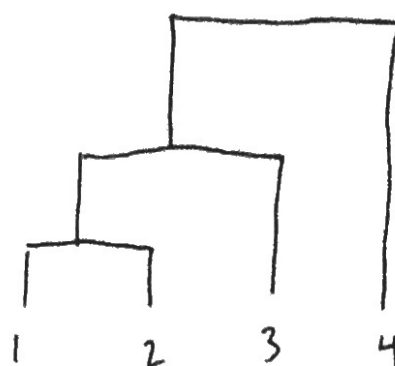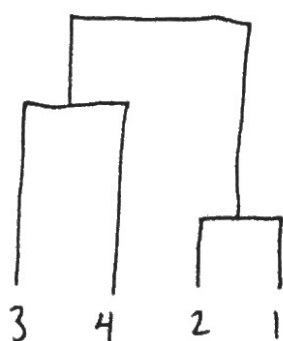2.a)

b)

c) 12, 34

d) 123, 4

e)

4. a) Not enough information. Points in these clusters may vary in how far they are from points in the other cluster, which can cause the dendograms to vary.

b) ~~Same height~~. Not enough information. Points other than 5,6 may cluster in different orders if, say, some are similar to point 3 and some aren't, but they're similar to each other.

6. a) The projections of each observation onto the first principal component are, on average, 10% of the length (in $\mathbb{R}^{100}$) of the observation vectors.

b) The preprocessing step in addition to controlling somewhat for machine choice, will do such an imperfect job at it as to be very inadvisable. Moreover a significant amount of the variation of the data itself is undoubtedly captured by the first principal component, and this information is being discarded. Finally, attempting to control for a binary effect such as this with a linear correction is not often smart.

Another possible approach would be to run two completely separate tests, one for machine A and one for machine B. Because there are 100 samples, there should be enough samples taken on each machine to provide reliable tests. If we want to use PCA (advisable since $p>n$), we should do two independent cluster analyses, one for the observations processed on machine A and the other for the observations processed on machine B.