1. LeCun's 1998 paper from last week and He's residual networks paper are both papers exploring a neural network architecture, but they are separated by more than 15 years. Keeping aside the depth of the network, how were the building blocks of the two networks different?

2. A ResNet 152 has lower error than a ResNet 34, but at the cost of more parameters. Under what conditions might you prefer the smaller model?

3. He and colleagues suggest that a residual connection might be useful " if the optimal function is closer to an identity mapping than to a zero mapping". Do you agree with their intuition? Can you think of an experiment that might validate this intuition?

4. The 152 layer ResNet has 36 residual blocks in the conv4 stage, but fewer than 10 for all other stages. why do you think this might be the case?