1. kNN methods are the real first non-parametric models that we have discussed in this class. Provide contrastive examples of cases where kNN would be preferred over linear models, like linear or logistic regression, <u>before</u> results are known. (In other words, it is easy to justify kNN when you have non-linearities in your data, but conceptually when would it be ideal to <u>start</u> with kNN approaches?)
2. The core of the kNN approach is the concept of data distances: data points that are closer together are assumed to be more likely to have come from the same underlying distribution. Provide an example where this assumption breaks down (i.e., when distance is not a good measure of similarity).