# Exercise 1

Exercise 1.b

Talk.politics.guns, talk.religion.misc report does not show a lower average distance when compared to itself vs. compared to different groups.

Rec.autos, comp.graphics are more diverse, for some metrics they shows the lowest average distance when compared to itself vs. compared to different groups, sometimes they shows the highest distance.

Comp.sys.mac.hardware documents show the lowest average distance among each other.

It is interesting that all document categories show small distances when compared to comp.sys.mac.hardware group. One can also see that the metric scores for different group comparisons appear to be correlated. That is the curves show qualitatively the same behaviour, for example the scores when comparing to comp.sys.mac.hardware drop with respect to talk.politics.guns. This is observed for all groups of documents across all metrics.

Exercise 1.c

What is meant by best separation?

Exercise 1.d

Since the Manhatten distance is simply the sum off each difference in the corresponding features we would expect the sum to approach infinity for infinitely dimensional datasets

For the Euclidean distance the

Since the dimensionality of our dataset is fixed we are not able to make a statement.