Assignment 1

1.

- For the Rothkopf Morse code data

The element in the matrix is defined as the percentage of subjects who think the two signals are the same. It can be viewed as a distance of similarity between two signals on the scale between 0 and 1. If we want the distance to be defined with the same direction. We just need to let element to be replace by (100 – original value) which is the dissimilarity between two signals. Then,

  - The positive and minimality still hold. The dissimilarity between two signals is always a positive number, with the upper bound of 1 and lower bound 0. If and only if all subjects believe the two signals are the same, the distance of dissimilarity is 0 and if and only if all subjects believe the two signals are different, the distance of dissimilarity is 1.
  - The symmetry is not hold. The data show that the judgements of subjects generally are not consistent (because judgement is subjective). Because this human-based noise, the symmetry is not strictly hold. For example, the distance of dissimilarity between A and B is 96 (100-4) but the distance of dissimilarity between B and A is 95 (100-5).
  - The triangle inequality is not hold. Since the symmetry is not hold, the distance cannot be unique defined. Or, the distance must hold the direction between starting point and ending point. One example that do not hold the triangle inequality is d (A, J), d (A, E), and d (E, J).

- For the Ekman's data

The color data has already been rescaled to the scale between 0 and 1. It can be viewed as a distance of similarity between two color on the scale between 0 and 1. If we want the distance to be defined with the same direction. We just need to let element to be replace by (1 – original value) which is the distance of dissimilarity between two colors.

  - The positivity and minimality still hold. The dissimilarity between two color is always a positive number, with the upper bound of 1 and lower bound 0. If and only if all subjects believe the two colors are the same (score as 4), the distance of dissimilarity is 0 and if and only if all subjects believe the two colors are different (score as 0), the distance of dissimilarity is 1.
  - The symmetry is hold. This is mainly because the pair of two color is only tested one time. Thus, we assume the symmetry hold directly.
  - The triangle inequality is not hold. This is many because of the human-based error after when compare so many colors. Their judgements can be subjective and inconsistent.

2. According to the axioms of a matrix space, d(a,b) for any arbitrary two points a and b must follows the triangle inequality. d(a,b) <= d(a,c) + d(c,b)

For four points x, w, y, and z, we can get:

- $d(x,z) \leq d(x,y) + d(y,z)$. If and only if x, z, and y are one the same line that equation fit.
- $d(x,w) \leq d(x,z) + d(z,w)$. If and only if x, z, and w are one the same line that equation fit.

Thus, we can get:

$d(x,w) \leq \underline{d(x,z)} + d(z,w) \leq \underline{d(x,y) + d(y,z)} + d(z,w)$

3. If we define the distance as the length of shortest path between two arbitrary points A and B. In other word, we can define a listable but usually countless set S which include the length of all path between A. Then distance is the minimun of the set S.

- for the minimality and positivity: as long as the path between A and B exist, the length of a path is always a positive number. If and only if A and B overlap together, the distance is 0.
- for symmetry: as long as the path between A and B exist, the paths (could be more than one path in this situation) with shortest length can always be identified. No matter how many paths there are, the smallest distance is unique. Consequence, the distance between A to B and B to A is still the same. It just may have multiple path to realize this distance.
- For triangle inequality: As long as the path between A and B exist. The only difference in this situation is that the path sometimes cannot be a straight line since the lacks exist. But it is always possible to separate the path into countable small straight lines. According to the problem 2, we have already proved that triangle inequality can be generalized to more than 3 points. The situation here is similar, for each separated straight line that make up the path, we only need to introduce one more point as the transition points. Using the same idea in the problem 2, when we sum all distance together, the triangle inequality still holds. If and only if the shortest path between A and B is not blocked by the lake, the equation fit.