**3. K-means**

**1.**

k = 2:

sum of within group sum of squares = 535116005.882

mistake rates = 0.522

k = 4:

sum of within group sum of squares = 459540485.768

mistake rates = 0.243

k = 6:

sum of within group sum of squares = 429366804.606

mistake rates = 0.18

**2.**

If k = 6, the number of iterations is 20, because convergence did not occur before the 20th iteration. Moreover, when I set the maximum number of iterations to a very large number instead of 20, the loop still did not terminate because of convergence, which means that when k = 6, convergence does not occur very early.

**3.**

**4.**