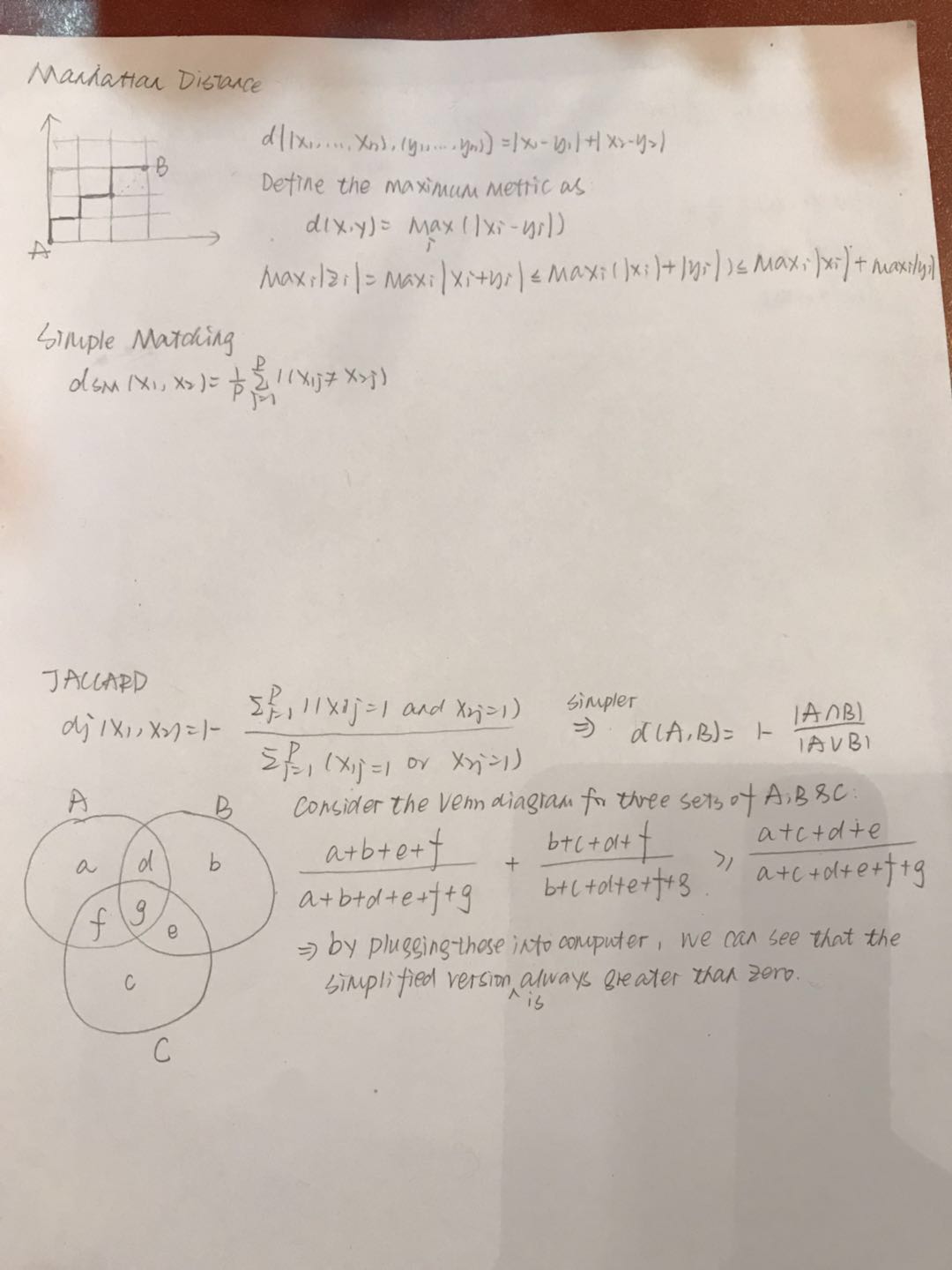
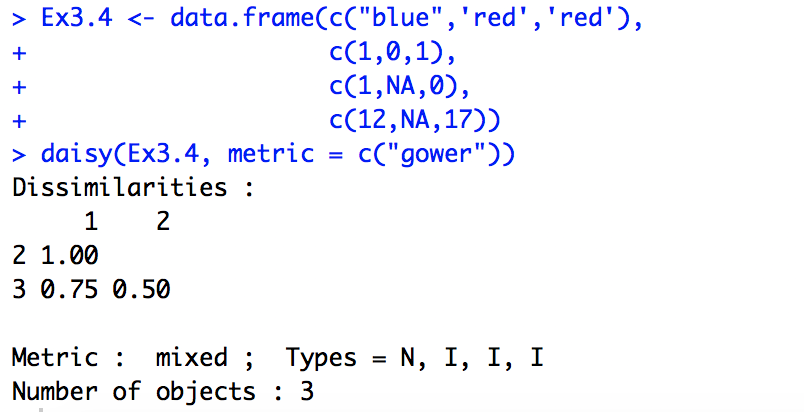
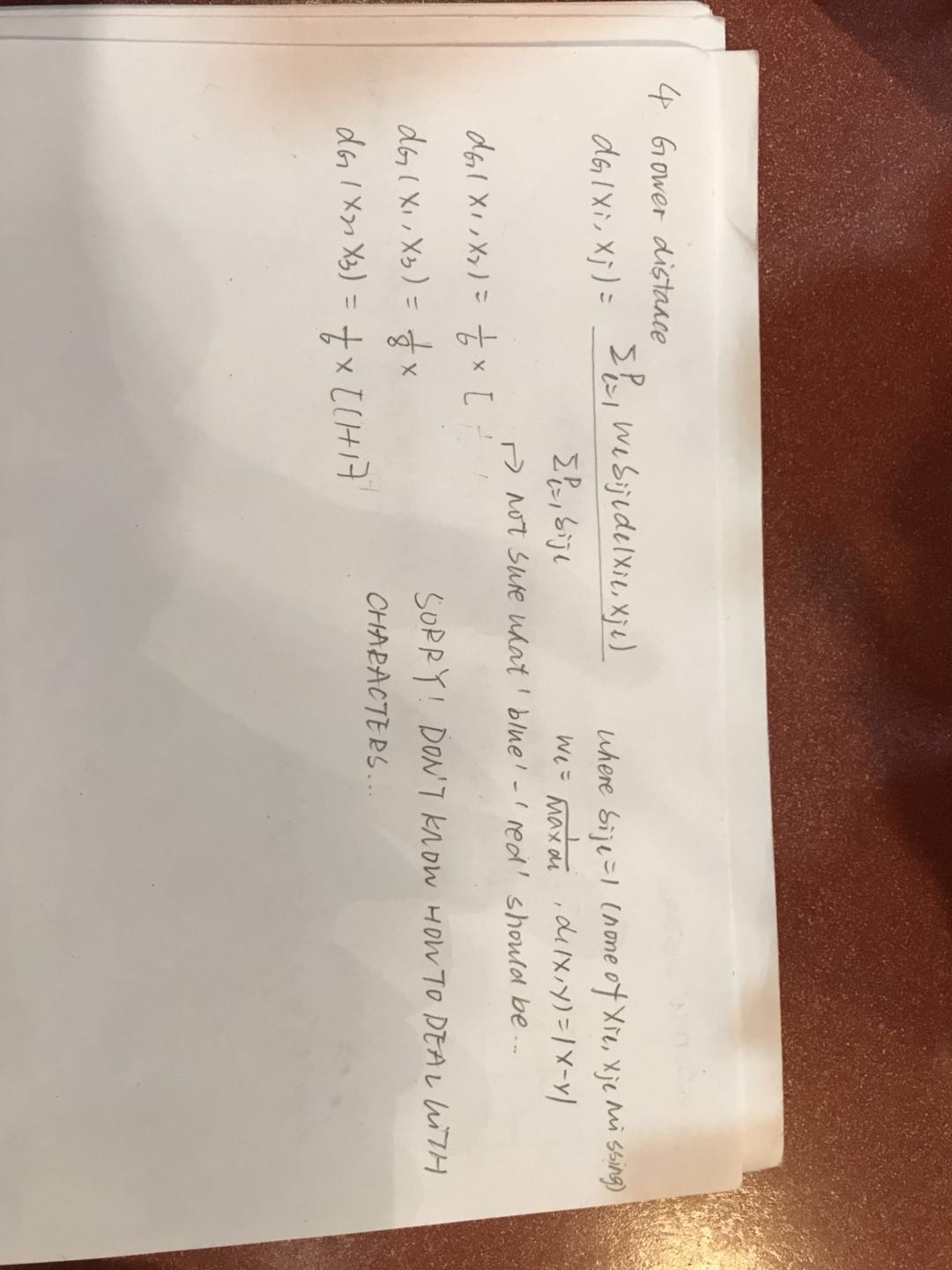
STAT 3019 Exercise 3 - Yina Lin



Q2: I am not quite sure about what this question is asking..

**Q4:**





Q5:

The Gap statistic standardizes the graph of log(Sk), where Sk is the within-cluster dispersion.

I did not know in which procedure of the clusGap should I remove the log. But I’ve read the paper ‘*A comparison of Gap statistic definitions with and with- out logarithm function’* written by M. Mohajer etc. and compared the plots that they produced. I saw that there are differneces between with and without logarithm. And reminded that the results are influenced by the trend, i.e. monotonic decreasing or increasing of the Gap function.

Conclusion of the paper: In fact, whenever a number fulfils the original Gap statistics inequality, this number also fulfils the inequality of a Gap statistics using the without log version, but not vice versa.