Statistics 503 Homework 6
Due: in class on Wednesday, Apr 27 2005

This homework is to be done individually, with help from Dr Cook as necessary. The length should be at most, 5 pages, excluding the appendix. It is worth 10 points, twice as much as the usual homeworks.

Background: This homework explores clustering of the movies that appeared in 2004. The data was extracted from the web site www.imdb.com. There are 63 movies and 4 measured variables (budget, length, average user rating, number of users rating). The data is available on the class web site.

Purpose: Learn about clustering methods, and to put together a case study without the help of your working group.

Exercises:
Put together a case study for this data, following the usual format. You need to:

• take the data description and plan for analysis from the web site, expand these as you see fit,
• conduct exploratory work
• conduct cluster analysis
• summarize your results and state your conclusions.
• reference appropriately