1. The data set I chose is “Epub” from the arules package. It contains the download history of documents from the Vienna University of Economics and Business Administration. When a user logs onto the platform a session is created, and the documents downloaded during that session are recorded.
2. The purpose of my analysis is to see what documents are frequently viewed together and possibly recommend or group together those documents in the future to cut down on search time for the user.

A screenshot of a computer

Description automatically generated with medium confidence

There were initially over a million rules, but I found many redundancies in the data and decided to greatly reduce the number of rules found to make the process more efficient for real world situations. For example, many permutations of multiple documents on the left-hand side, when one of those documents alone would satisfy the requirements for the right-hand side, were eliminated. For the sake of real-world use and cutting back on time searching for these “recommended” documents in the platform, I also chose to lower confidence to help keep the most frequent rules in play while also slightly increasing support and increasing coverage.

1. The rules that are most interesting to me are the ones with 5 or more documents on the lhs and predict with a confidence of 1 that the user will download the rhs. I think these are users who try to cover all their bases when looking up documents and download everything they can find on a certain subject. These users also help inform on the data itself and can help make stronger recommendations.