The file “TVShows.csv” contains data on which TV shows a random sample of 2000 subscribers to a paid streaming service have watched. Each row represents one subscriber. A value of 1 for a TV show indicates that the subscriber has watched at least six episodes of that show; otherwise, the value is 0. The variables are defined this way in an attempt to capture whether a subscriber enjoys the show (or at least is willing to watch it regularly).

The streaming service provider sometimes displays advertisements for shows, which can be targeted strategically to subscribers who are more likely to be interested in them based on their viewing history. You are the team of analysts working for the provider, and you have been tasked with improving strategic targeting of the ads by using association rules.

You will need to start by importing the TVShows dataset into R.

Each group will be asked to answer **one** of the following questions and explain their answer to the rest of the class:

**A**: Your company has agreed to display a promotional ad for *Archer* 10000 times. To which subscribers should this ad be shown, and why? (As a reminder, the 2000 people in the dataset are a sample, and comprise only a very tiny fraction of the overall number of subscribers.)

{Ash vs. Evil Dead}

Because they have the highest life ratio

**B**: The most popularly watched show is *Game of Thrones*. What show(s) would you recommend promoting to the subscribers who watch *Game of Thrones*, and why? (As a reminder, the 2000 people in the dataset are a sample, and comprise only a very tiny fraction of the overall number of subscribers.)

**C**: The show *You’re the Worst* has not been as popular as your company had hoped, and they would like to promote it more aggressively. To which subscribers would you recommend promoting *You’re the Worst*, and why? (As a reminder, the 2000 people in the dataset are a sample, and comprise only a very tiny fraction of the overall number of subscribers.)

**D**: Currently, your company is advertising only their most popular shows, such as *Game of Thrones* and *House of Cards*. However, based on recent customer focus groups, they have determined that the ads subscribers find the most valuable are ads for shows that aren’t quite as popular, but are tailored specifically for their tastes. In light of this finding, they have asked you to come up with a recommendation for which additional show(s) to advertise, and to which subscribers.

**E**: Your company’s current ad targeting strategy is as follows: when a subscriber watches a show, display an ad for whatever other show is the most popular among viewers of the show being watched. What is the weakness of this strategy, and how would you recommend improving it? Provide a specific example.

**Tip**: The “sort” function used in the Association Rules R script can only sort based on the numeric columns. Usually that’s what you want, but it might also be helpful to sort based on lhs or rhs. There are multiple ways to do that, but the following lines of code are the most straightforward:

rules.sorted <- rules[order(labels(lhs(rules)))]  
rules.sorted <- rules[order(labels(rhs(rules)))]

**Another Tip:** There is a function called “subset” that can do more elaborate kinds of filtering on sets of association rules: <https://www.rdocumentation.org/packages/arules/versions/1.6-8/topics/subset>

If you finish early, feel free to try it out. Here’s one example of something you could do with it:

rules.subset <- subset(rules, subset = lhs %in% "House of Cards" & lift > 1.1)

Then inspect rules.subset and see what you get. (This example is just for demonstration purposes; it doesn’t address any of questions A through E above.)