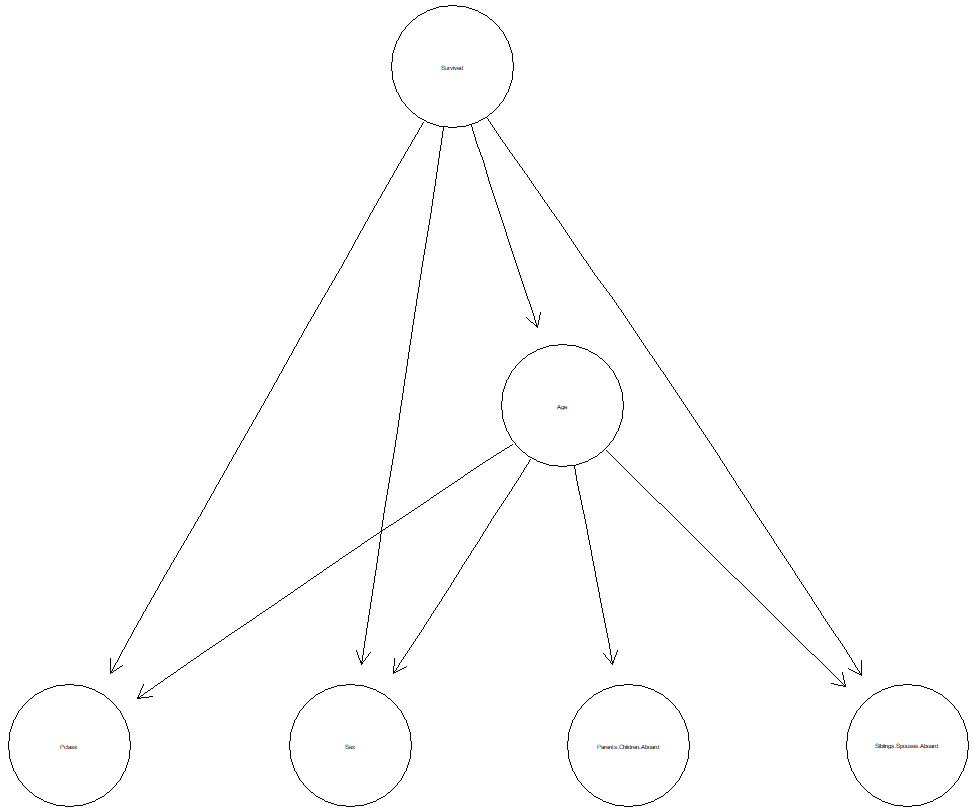
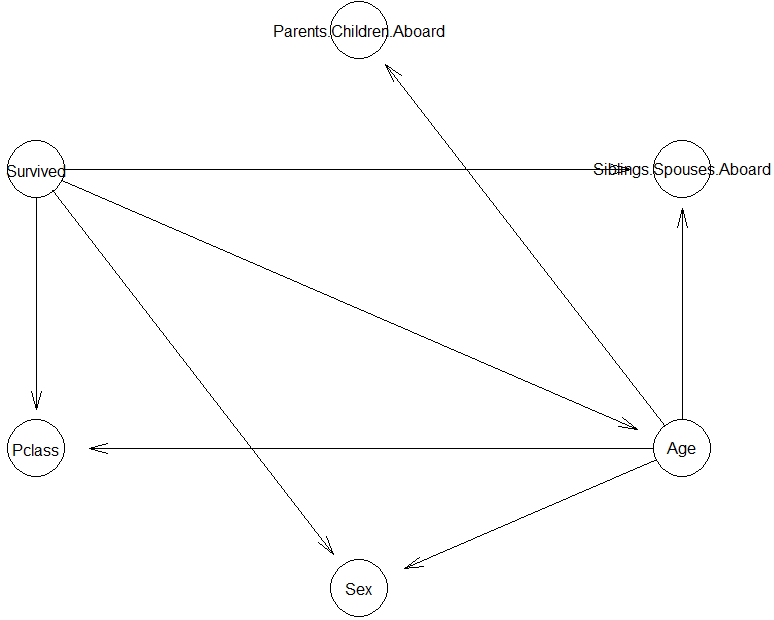
Question 2

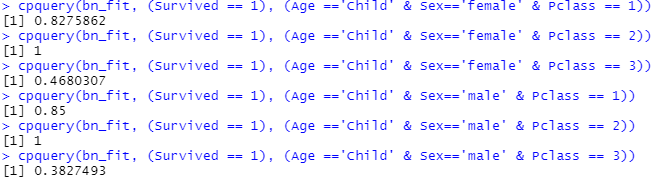


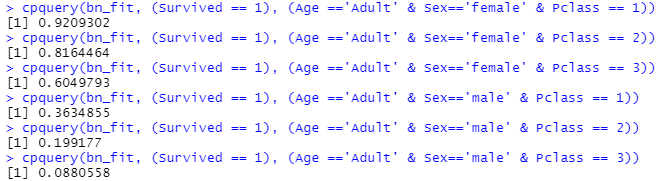
1. Prob of women and child survival



Probability of women and child survival is about 0.70 and 0.5. Let’s include more conditions below to see the detailed probability.

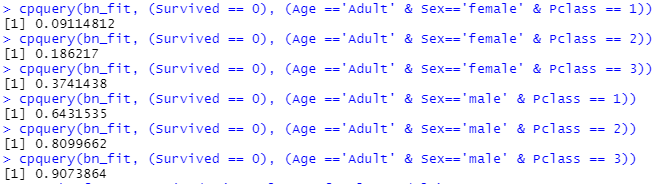
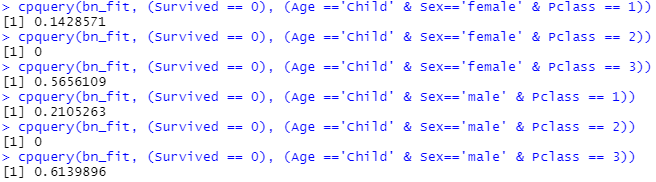
1. What characteristics/demographics are more likely in surviving passengers?





As you can see, all of the 2nd Pclass children survived 100%. Besides that, the highest possibility of survival is when a person is in 1st class, female and adult. On the other hand, the lowest possibility of survival is when a person is in 3rd class, male and adult.

1. What characteristics/demographics are more likely in passengers that perished?



As you can see, the chances of survival is very low when a person is in 3rd class, adult and male.

1. Prob of Rose survived, (1st class & female & Adult)



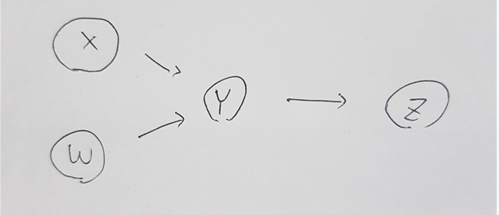
We can conclude that since Rose is female, 1st class passenger, and adult, the possibility of survival is very high, which makes sense because she survived in the movie.

1. Prob of Jack died, (3rd class & male & adult)



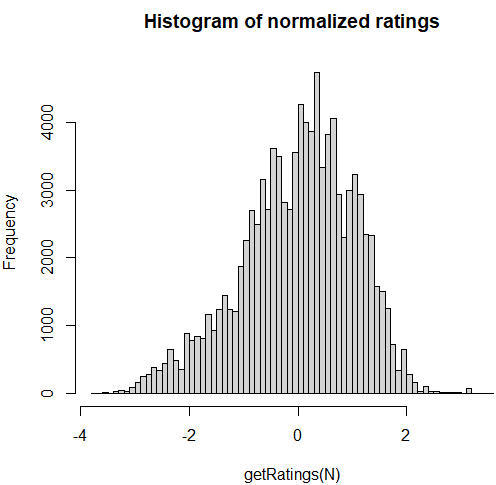
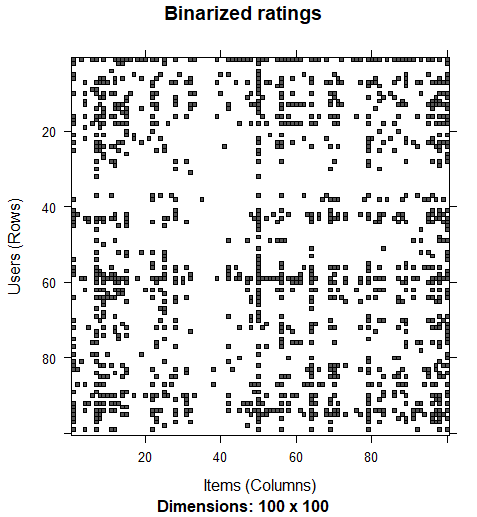
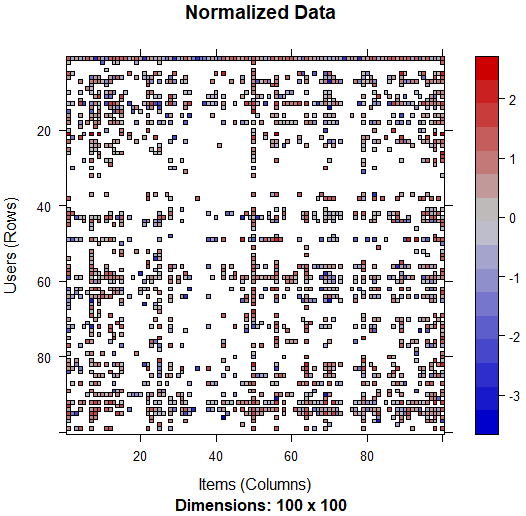
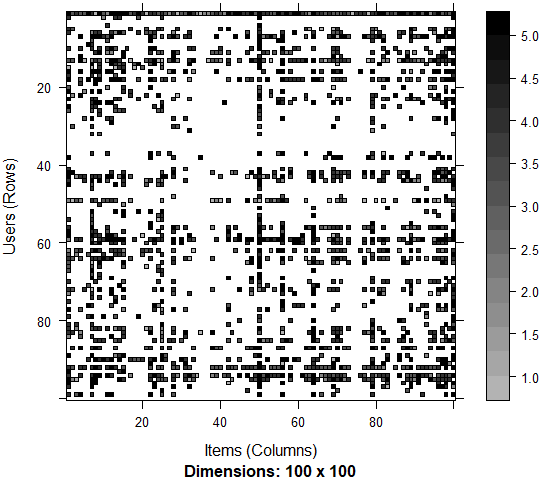
We can conclude that since Jack is male, 3rd class passenger and adult, the possibility of survival is very low, which makes sense because he died in the movie.

Question 3

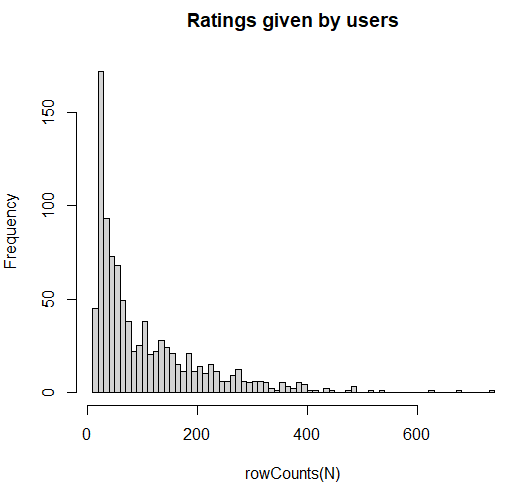


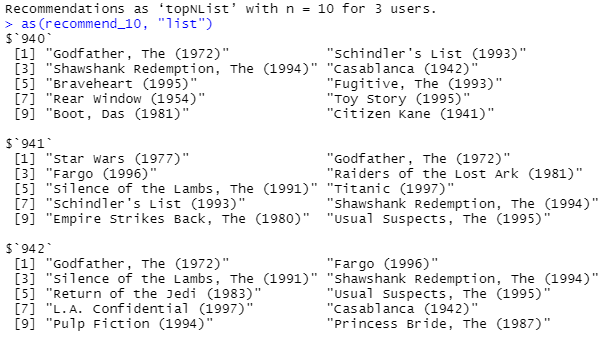
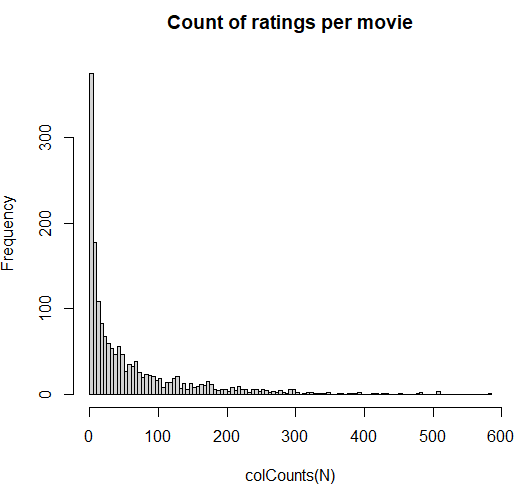
Question 4

Below is 100X100 image of MovieLense data rating

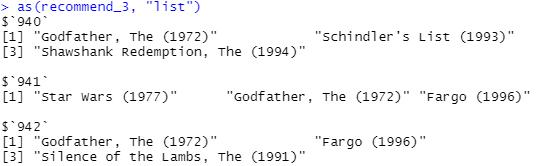


Most ratings are in the zero range, which tells that people do not give high ratings very often.



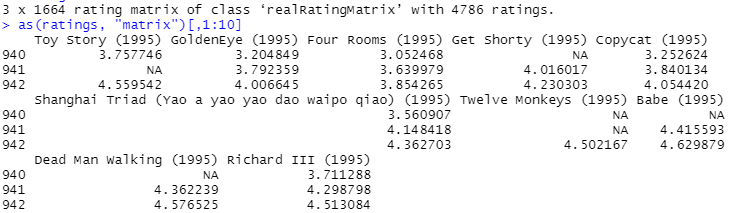


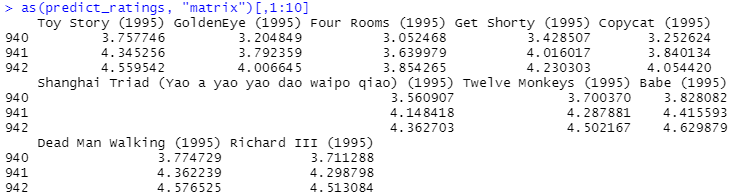
I picked random 3 users who rated the top movie ratings. Similar movies such as Godfather, toy story, fargo and etc can be seen from the data.



Among the top 3 recommendations, godfather appeared the most.

Predicted ratings with NA and without NA is shown below.

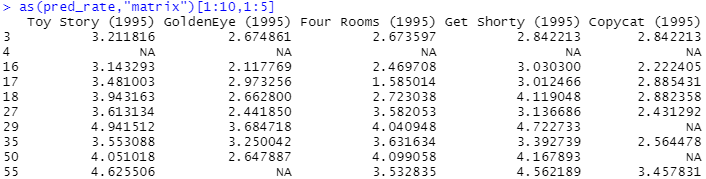




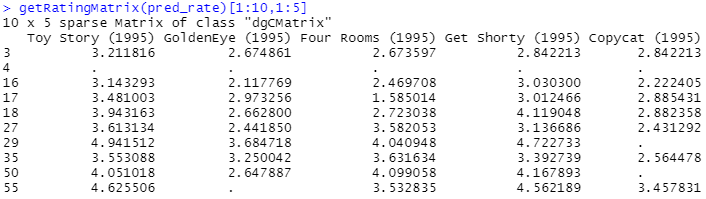


Question 5

Cross-validation is used.



This is the predicted rate of 10 users for 5 movies.



Predicted error is as follows.



Compared to question 4, question 5’s RMSE, MSE, and MAE are all higher. This means that they are not performing well.

We do 5 folds and average the result

