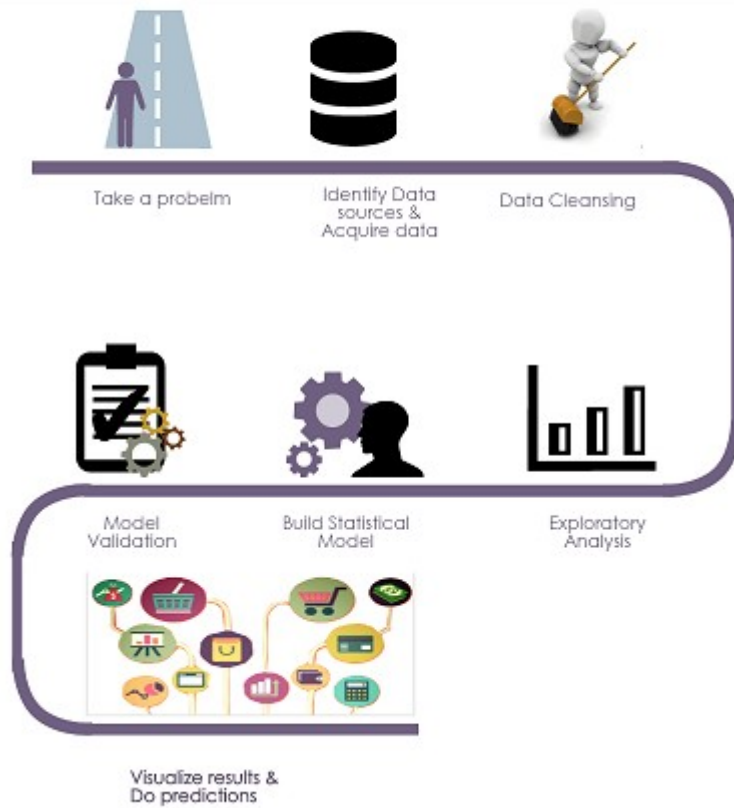


Chapter 2, Data Mining Techniques used in Recommender Systems

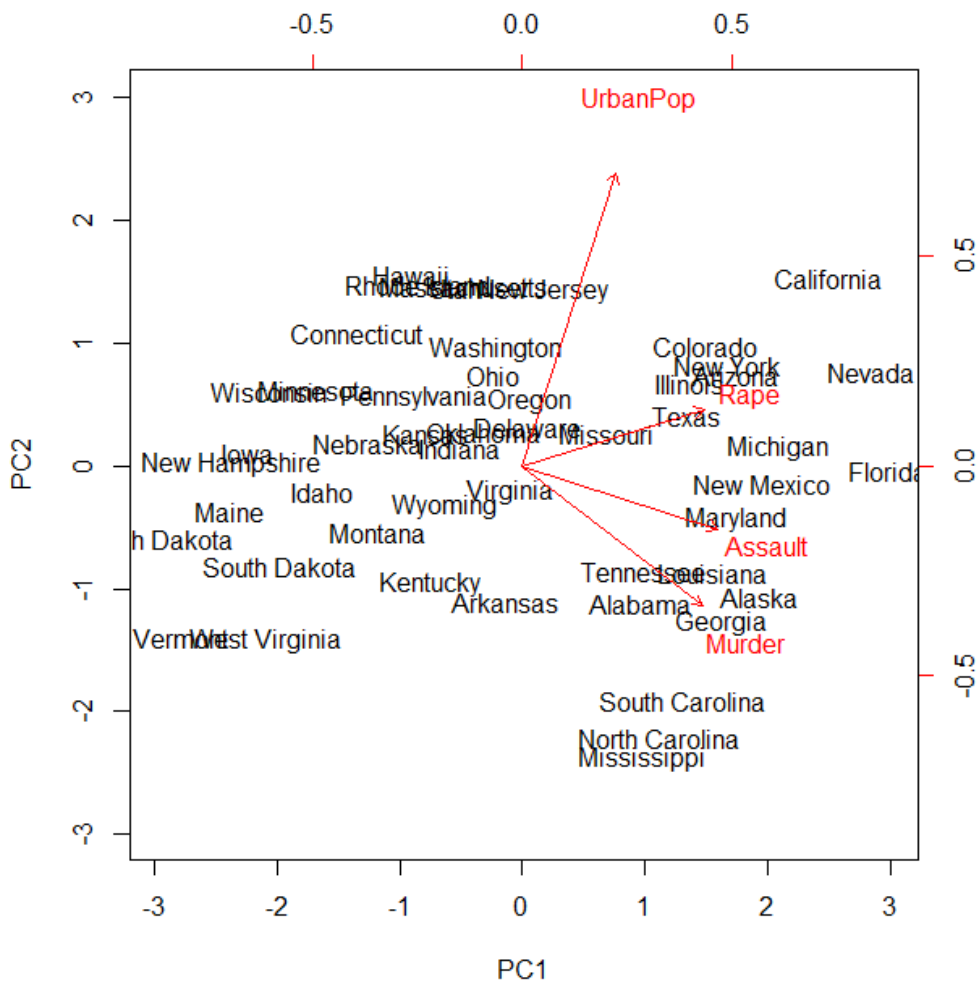
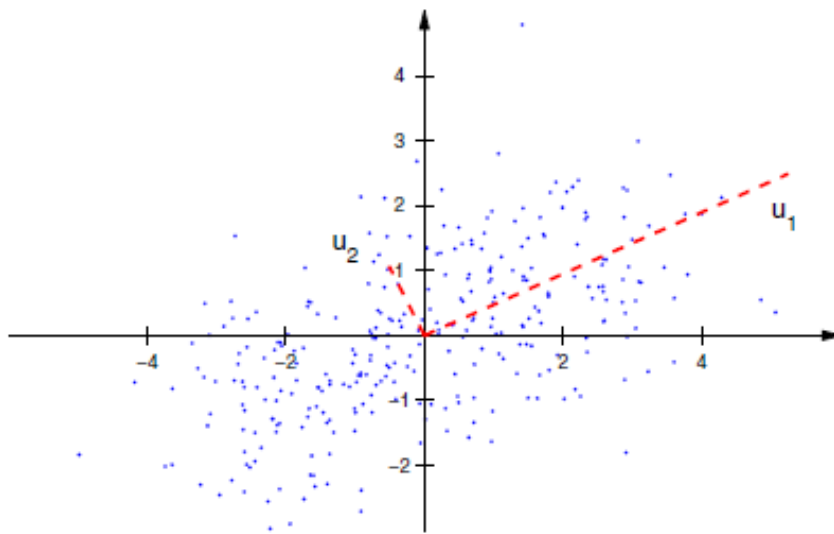
How to solve a data analysis problem

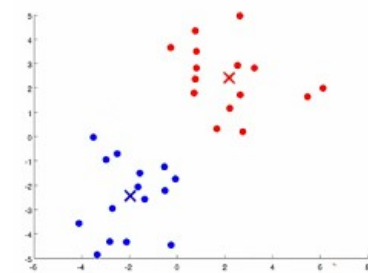
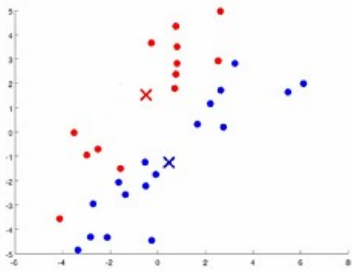
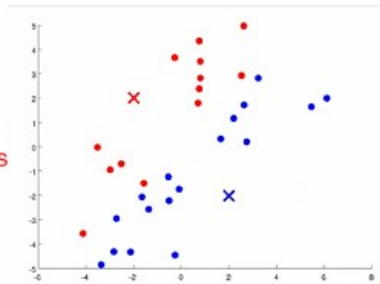
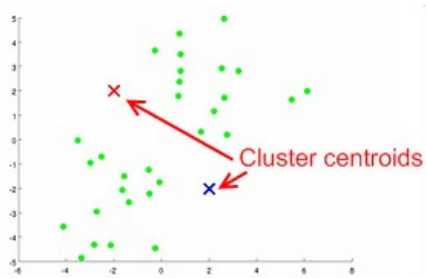
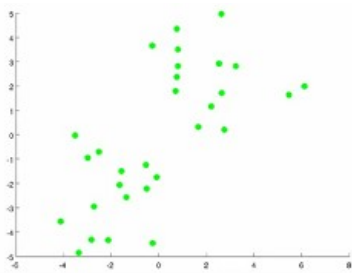
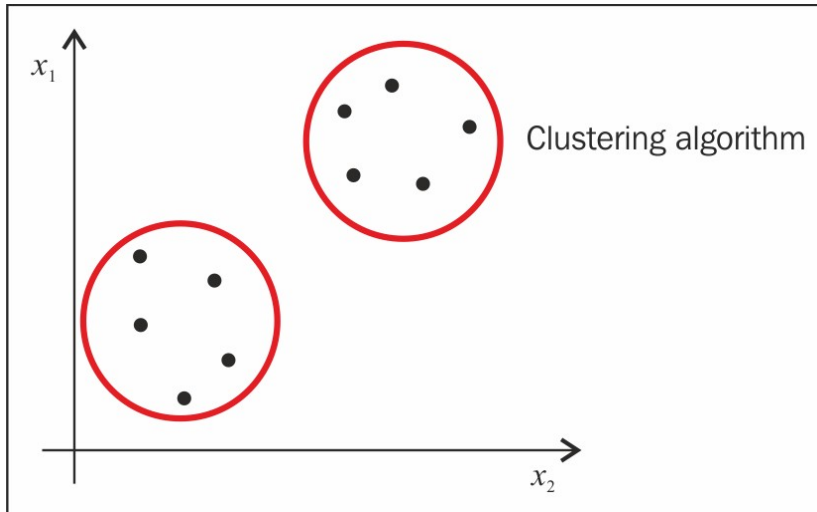


$$\text{Euclidean Distance}(x,y) = \sqrt{\sum_{i=1}^n |x_i - y_i|^2}$$

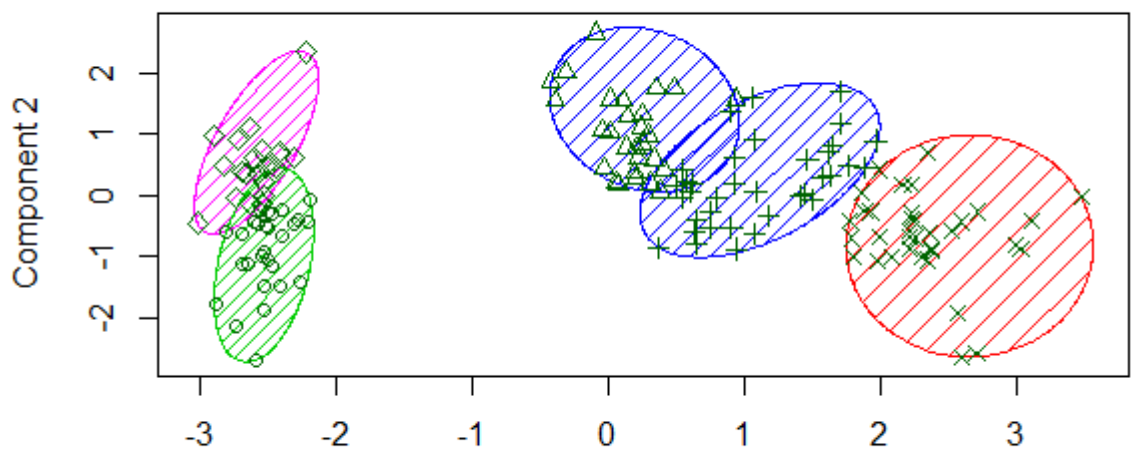
$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}$$



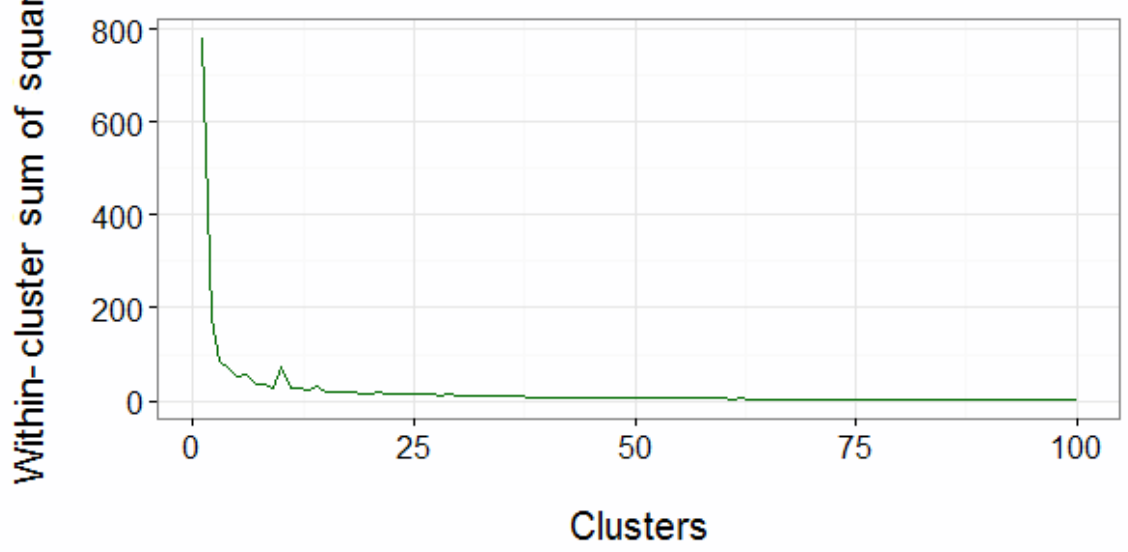


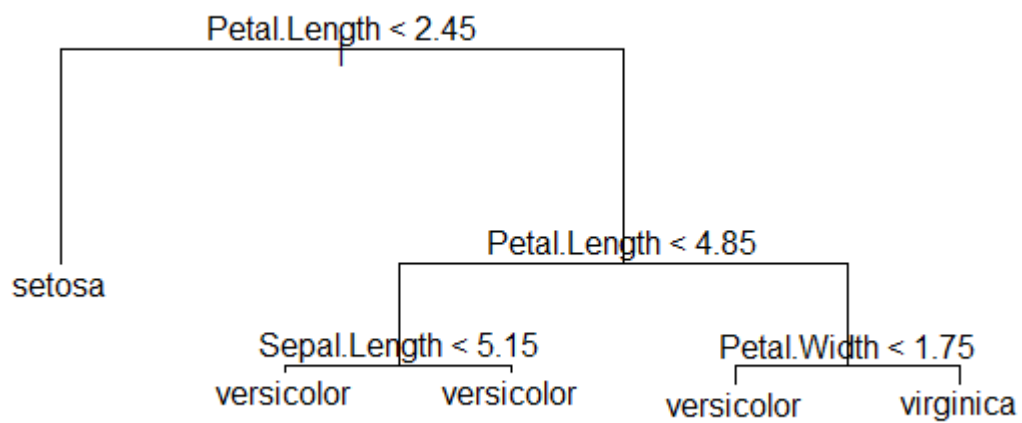
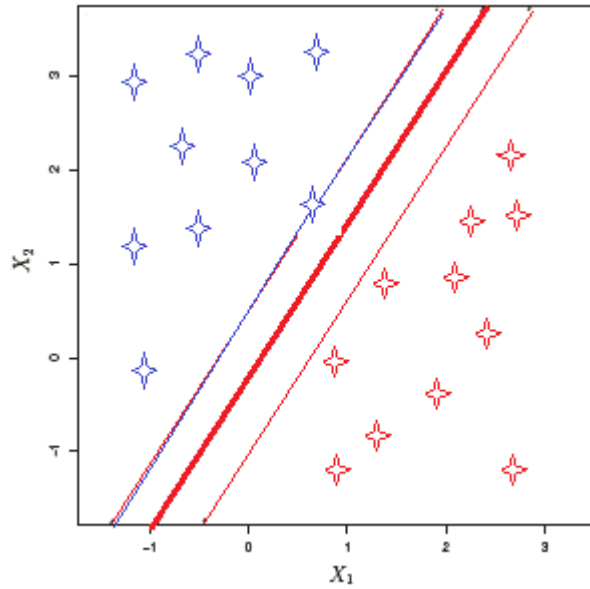
CLUSPLOT(iris)



Component 1
These two components explain 95.02 % of the point variability.

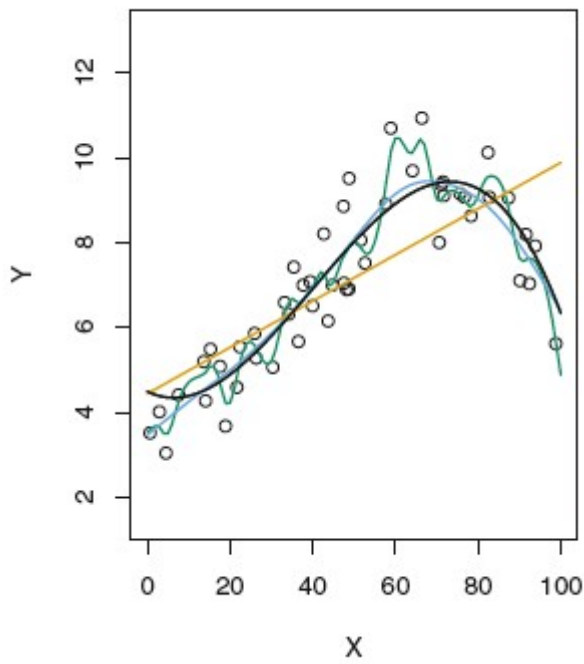
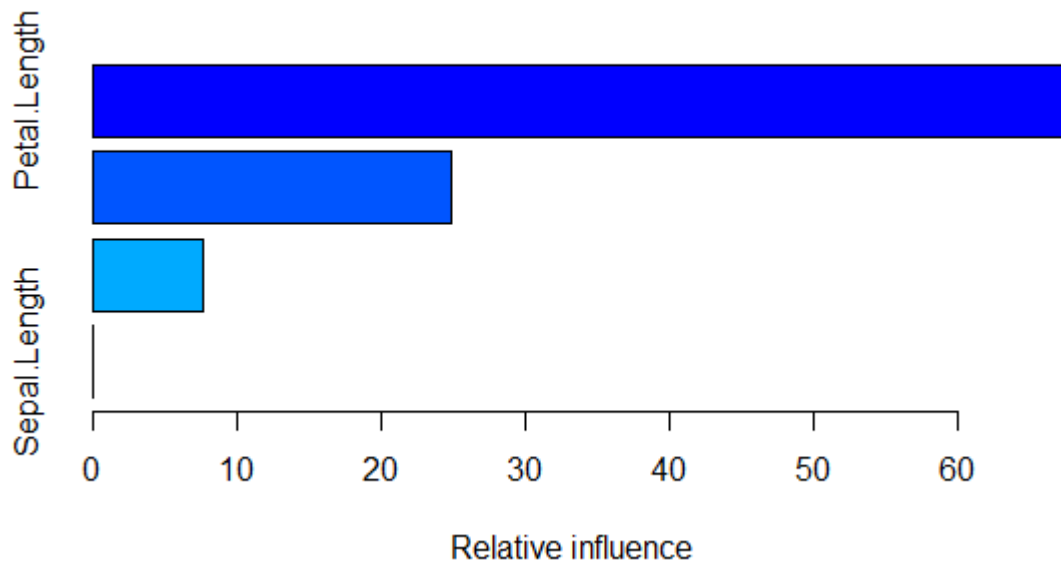
Reduction in cost for values of 'k'





```

> summary(model)
              var  rel.inf
Petal.Length Petal.Length 67.440852
Petal.width  Petal.width  24.942084
Sepal.width   Sepal.width   7.617065
Sepal.Length Sepal.Length   0.000000
  
```



		ACTUAL	
		POSITIVE	NEGATIVE
PREDICTED	POSITIVE	<i>TRUE POSITIVE</i>	<i>FALSE POSITIVE</i>
	NEGATIVE	<i>FALSE NEGATIVE</i>	<i>TRUE NEGATIVE</i>

sensitivity or true positive rate (TPR)

$$TPR = TP/P = TP/(TP + FN)$$

specificity (SPC) or true negative rate (TNR)

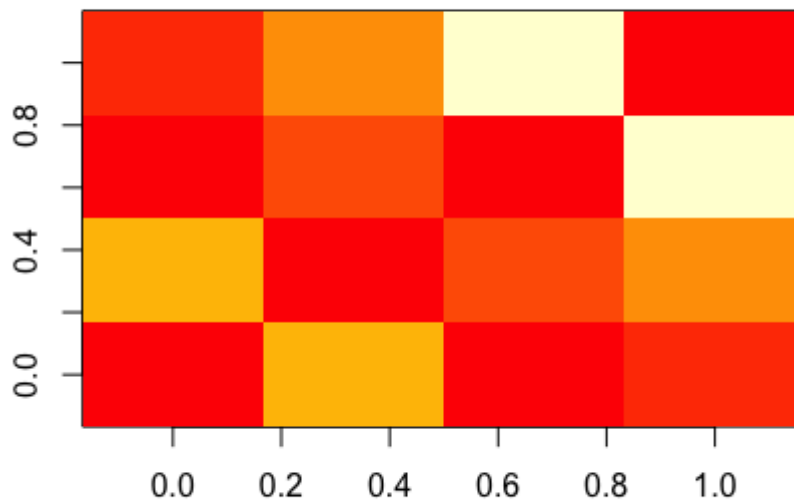
$$SPC = TN/N = TN/(TN + FP)$$

precision or positive predictive value (PPV)

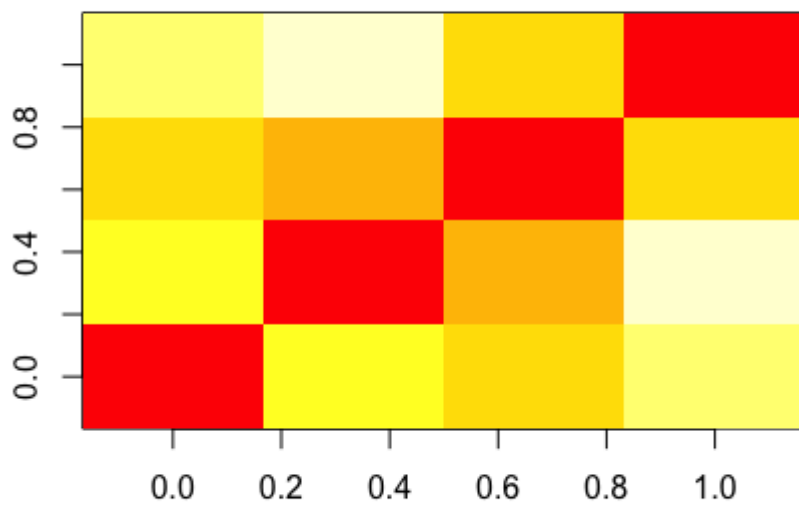
$$PPV = TP/(TP + FP)$$

Chapter 3, Recommender Systems

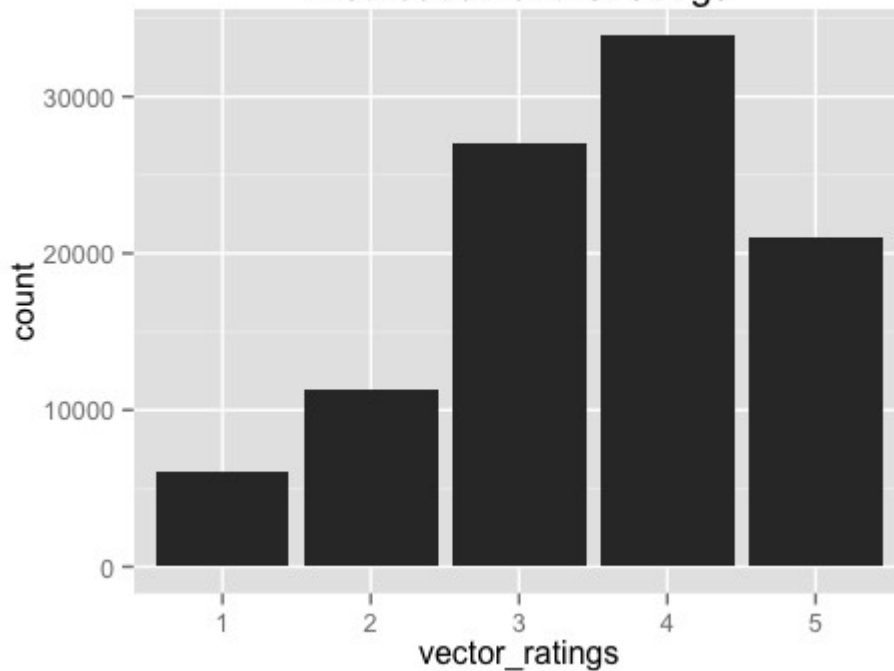
User similarity

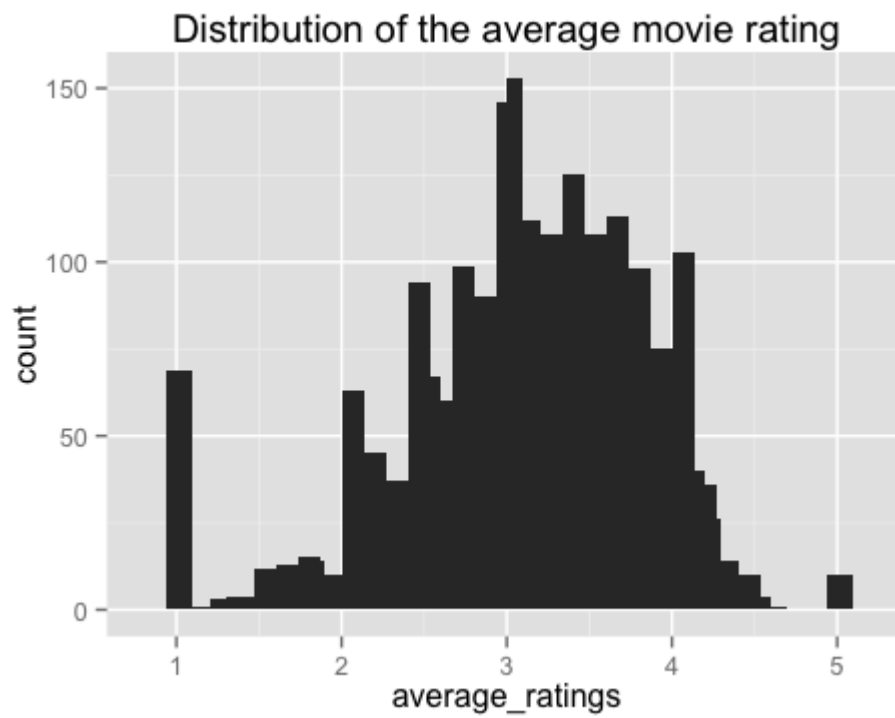
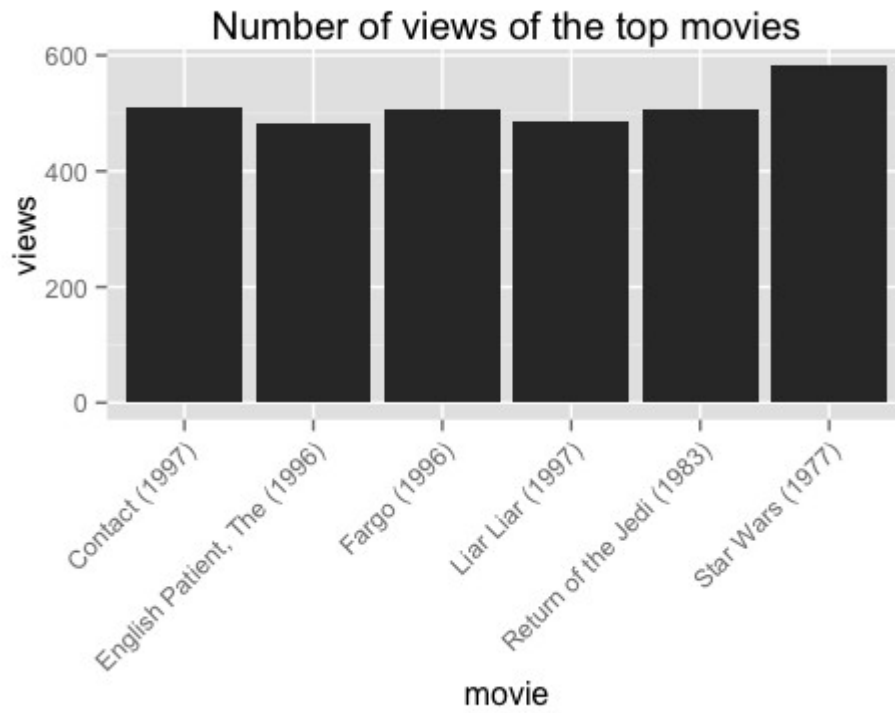


Item similarity

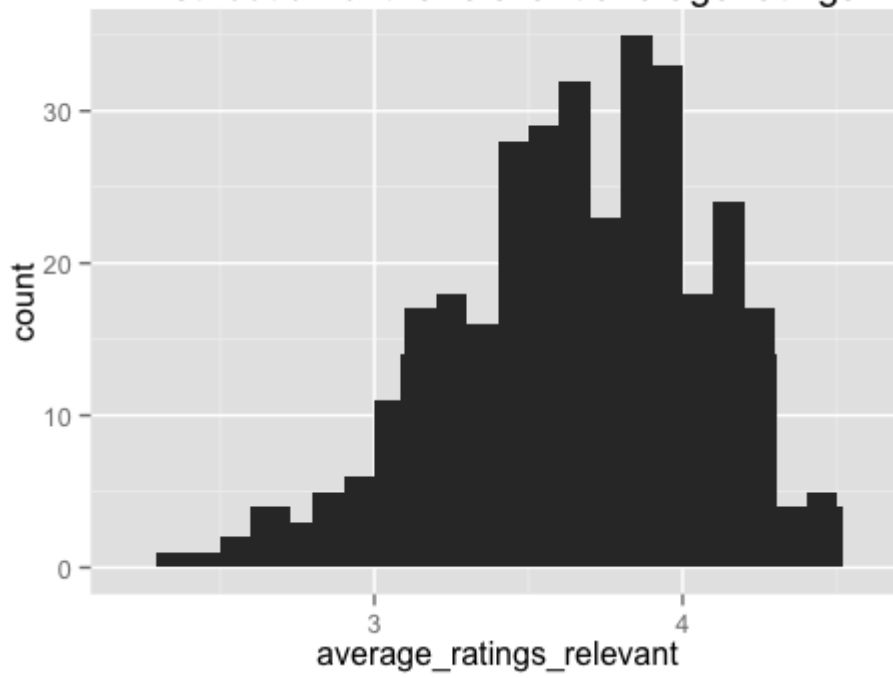


Distribution of the ratings

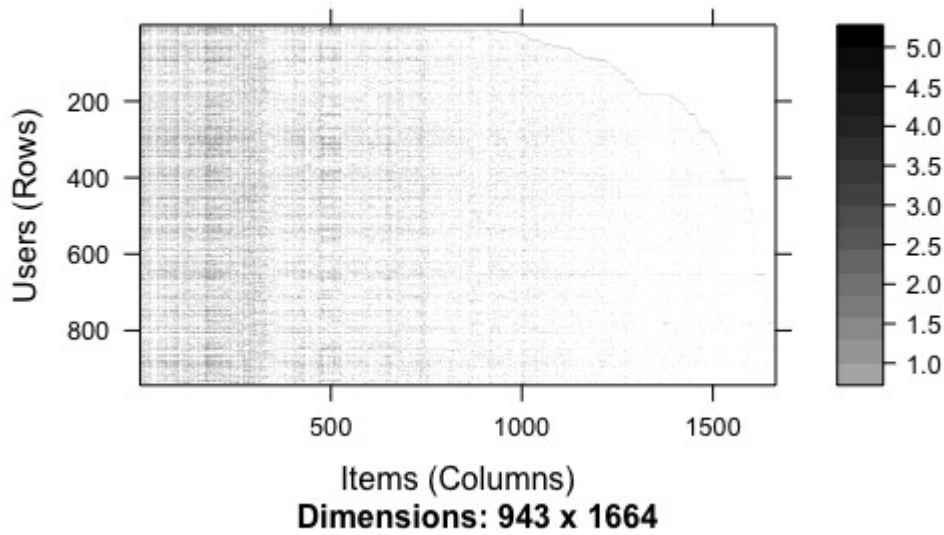




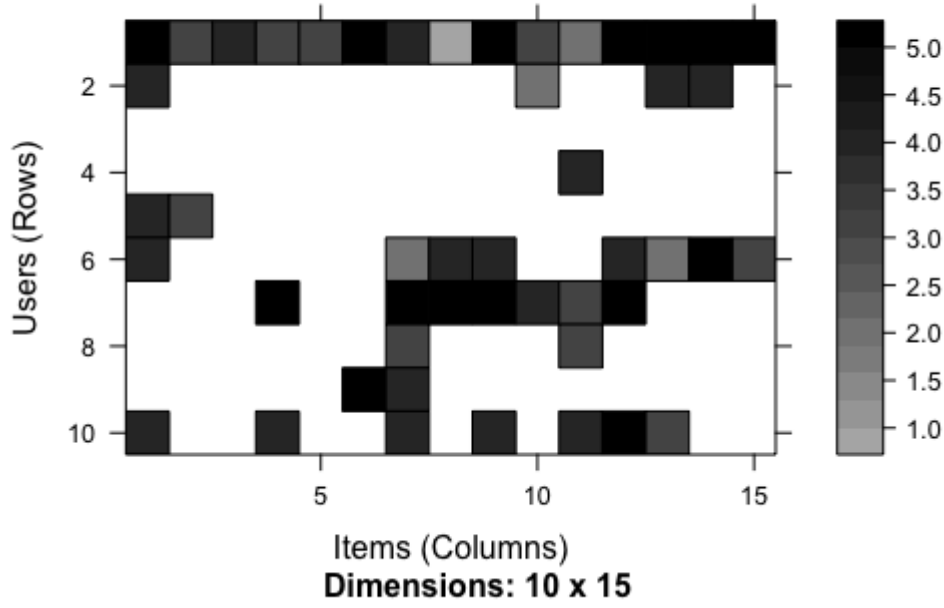
Distribution of the relevant average ratings



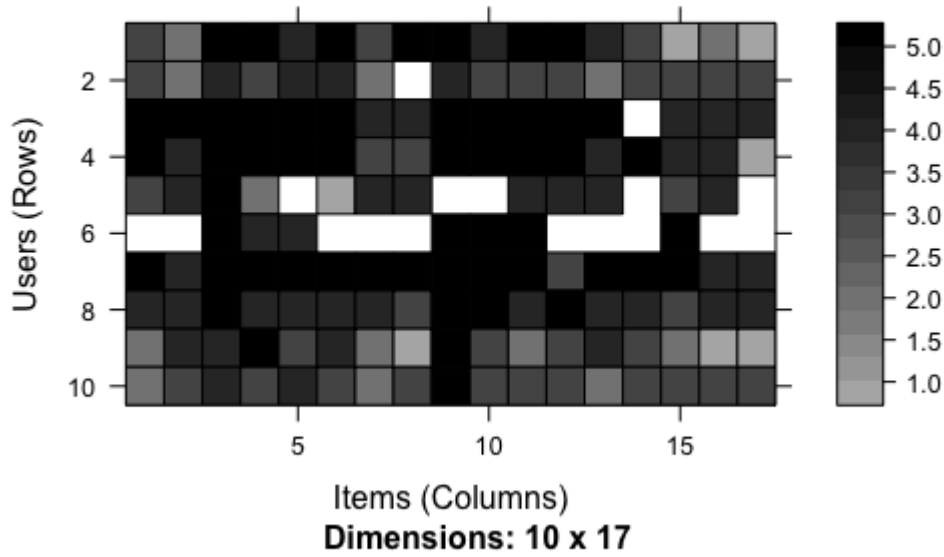
Heatmap of the rating matrix



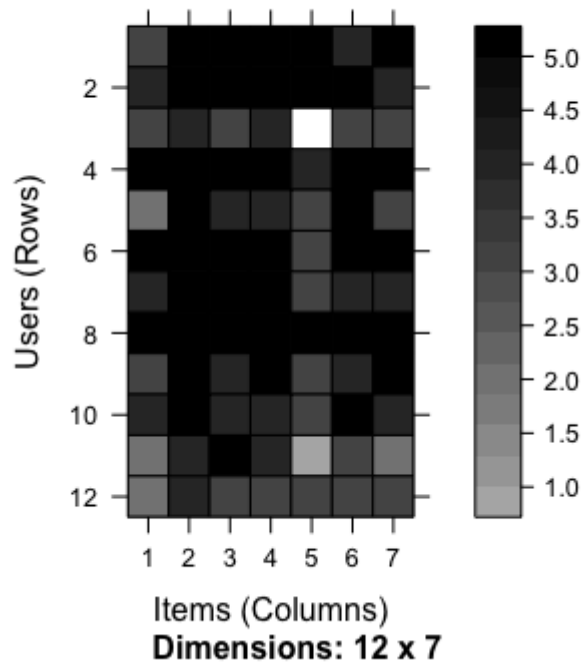
Heatmap of the first rows and columns



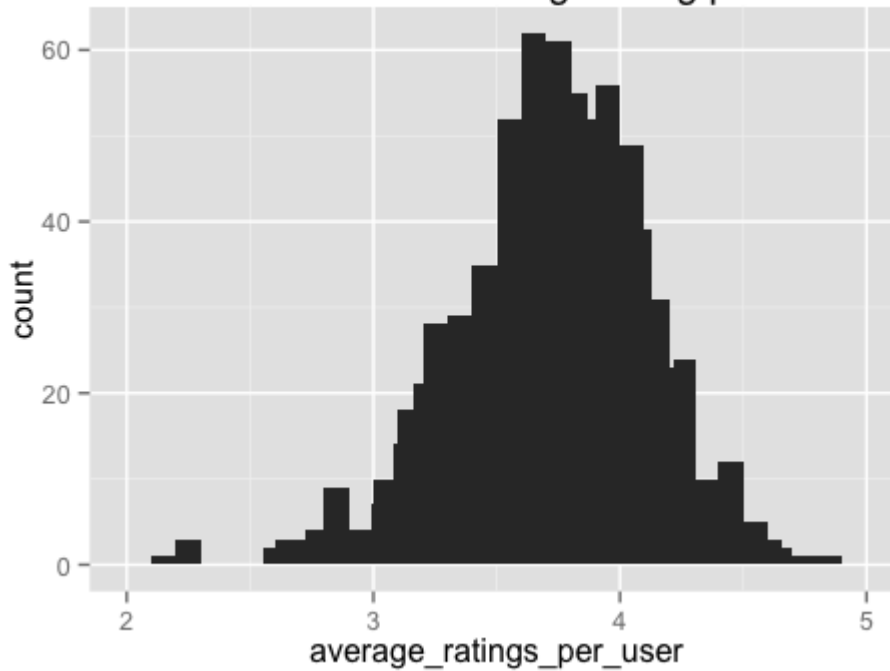
Heatmap of the top users and movies



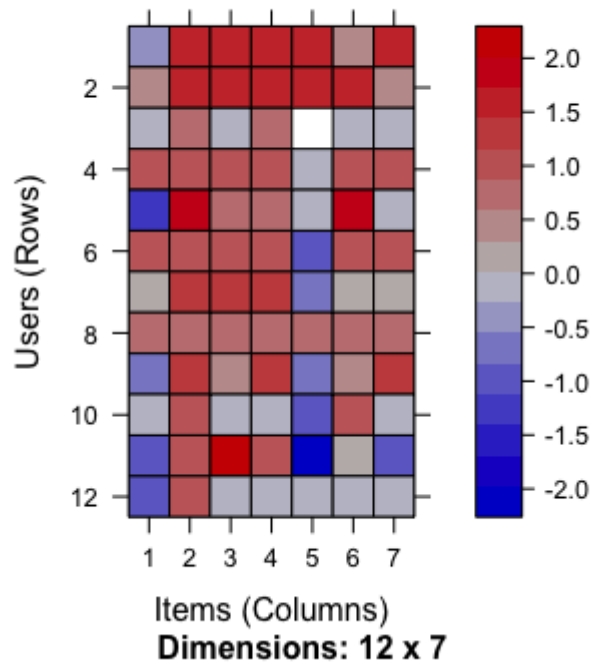
Heatmap of the top users and movies



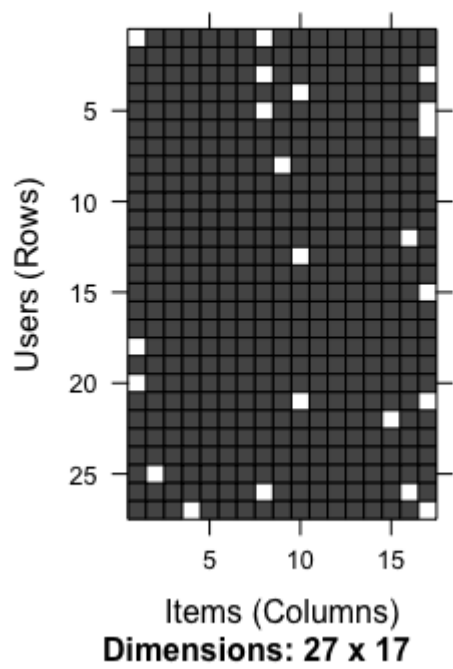
Distribution of the average rating per user



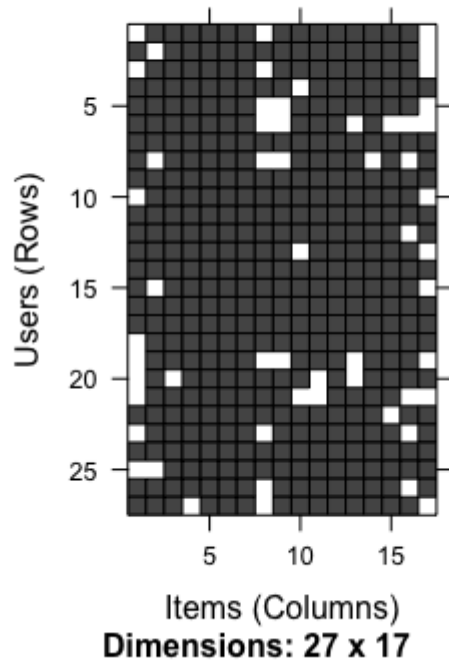
Heatmap of the top users and movies



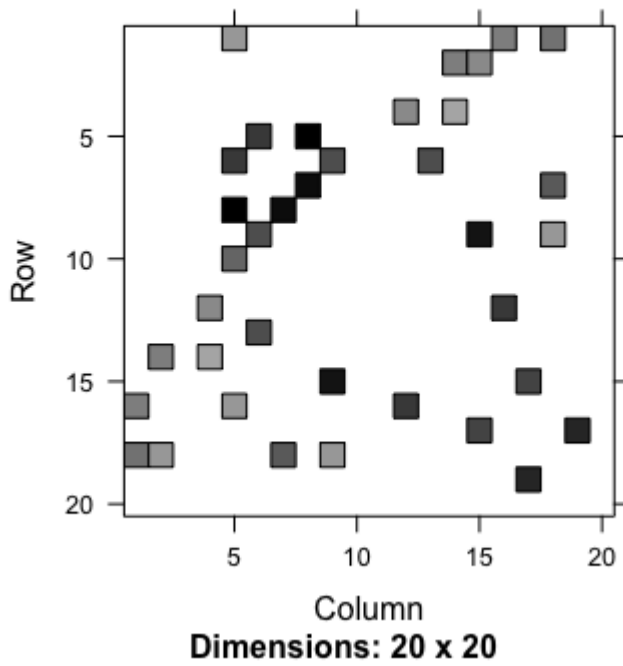
Heatmap of the top users and movies



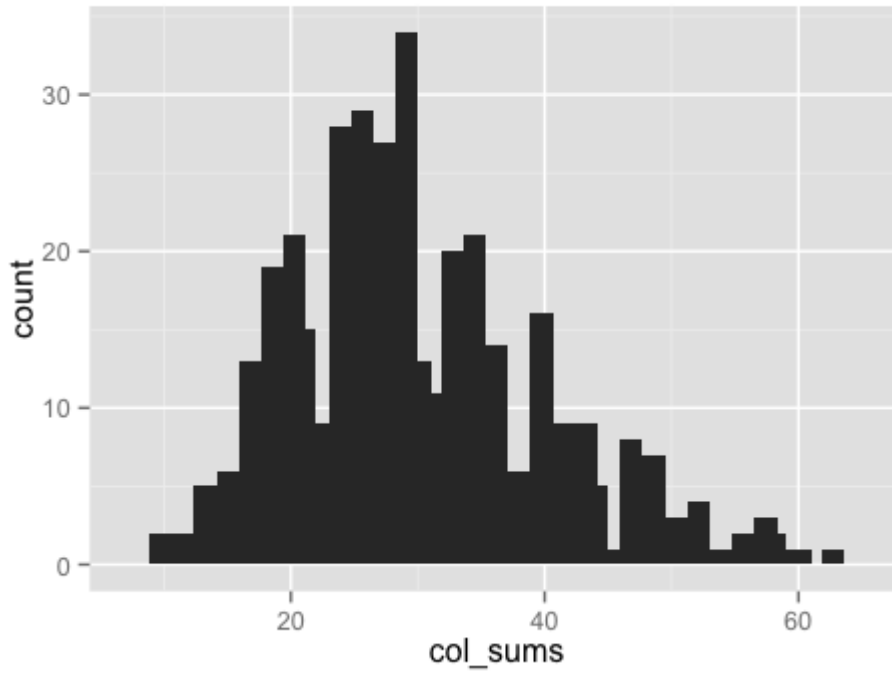
Heatmap of the top users and movies



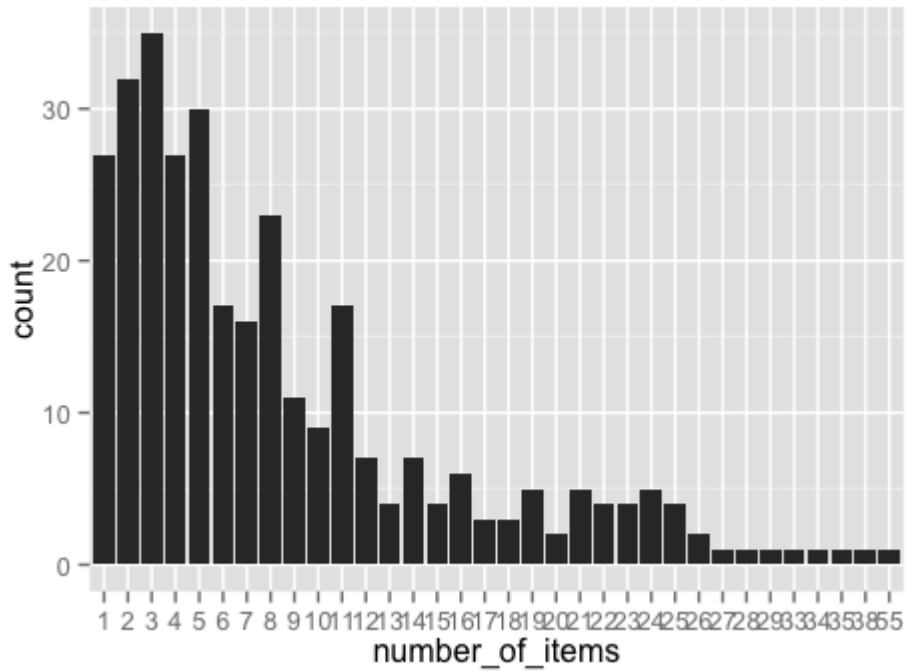
Heatmap of the first rows and columns

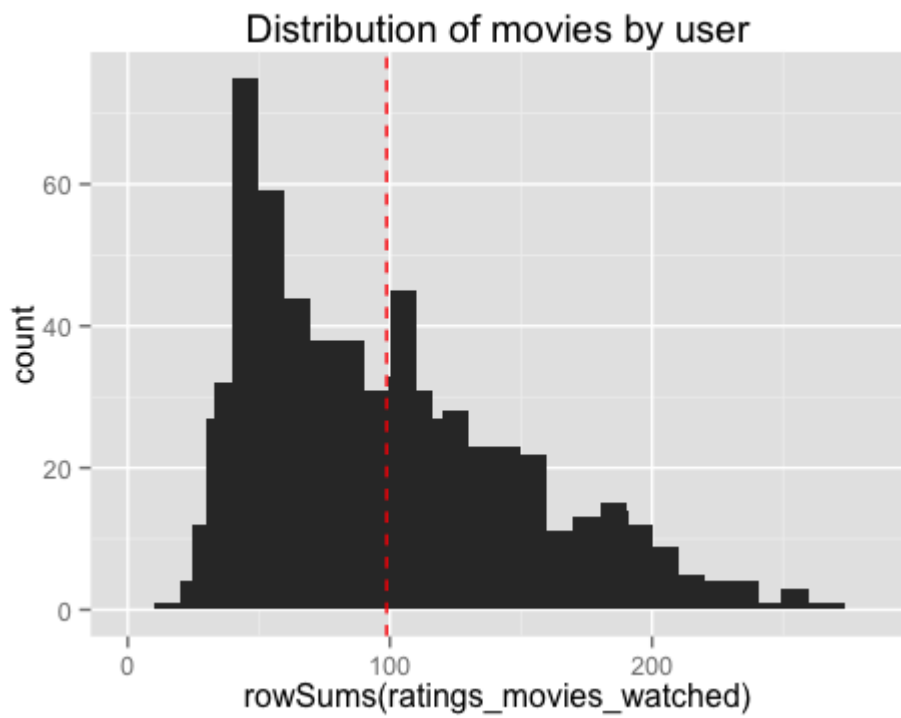
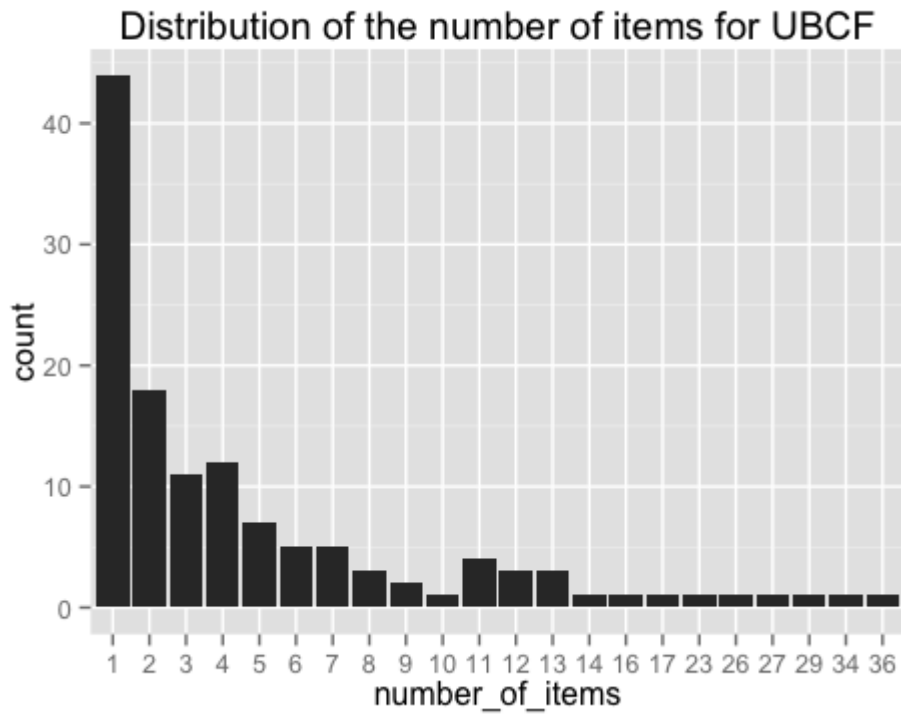


Distribution of the column count



Distribution of the number of items for IBCF





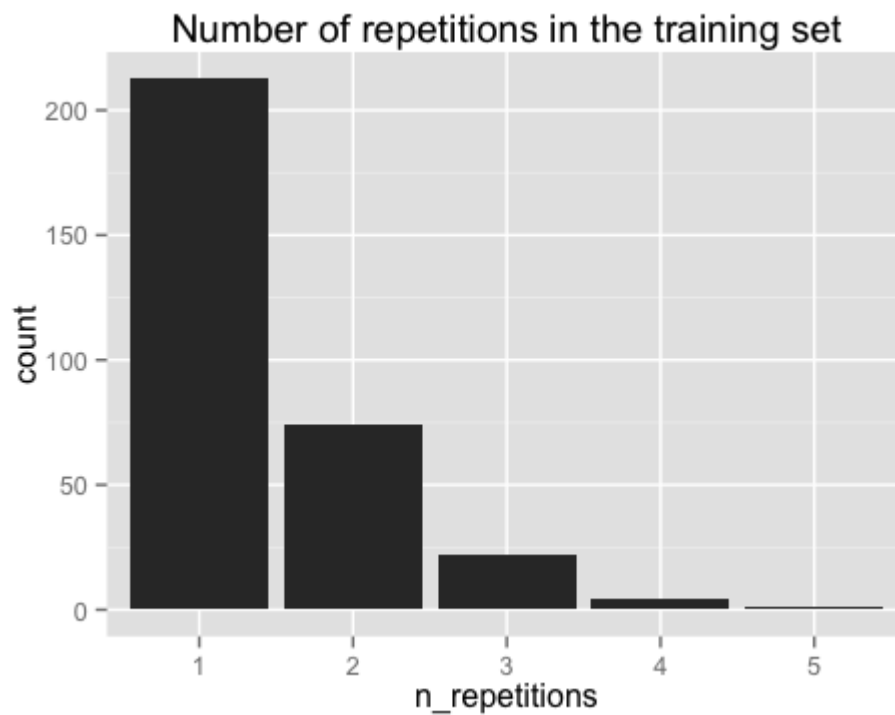
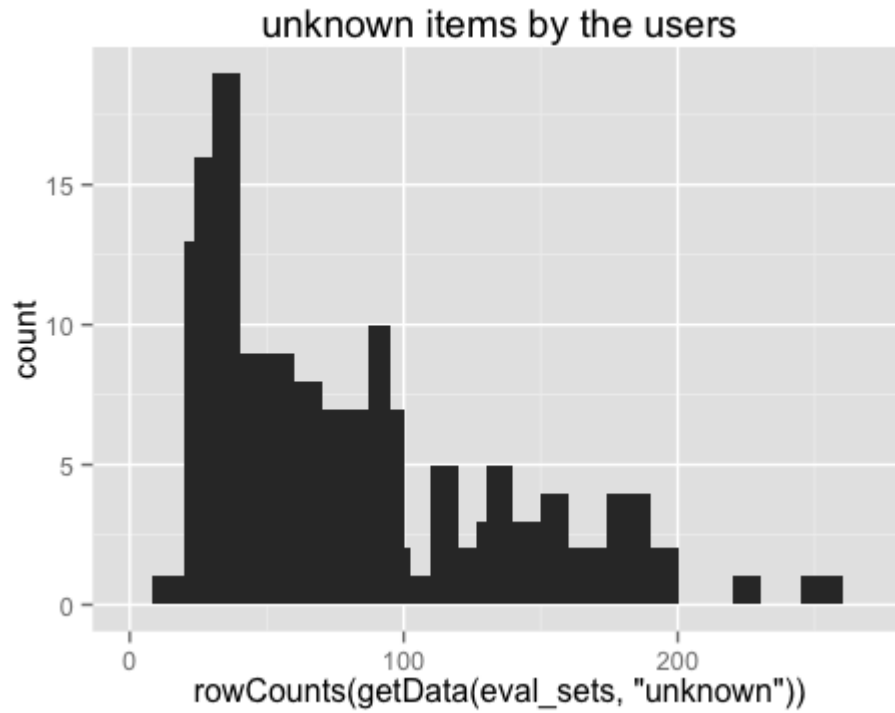
$item_1$

$item_2$

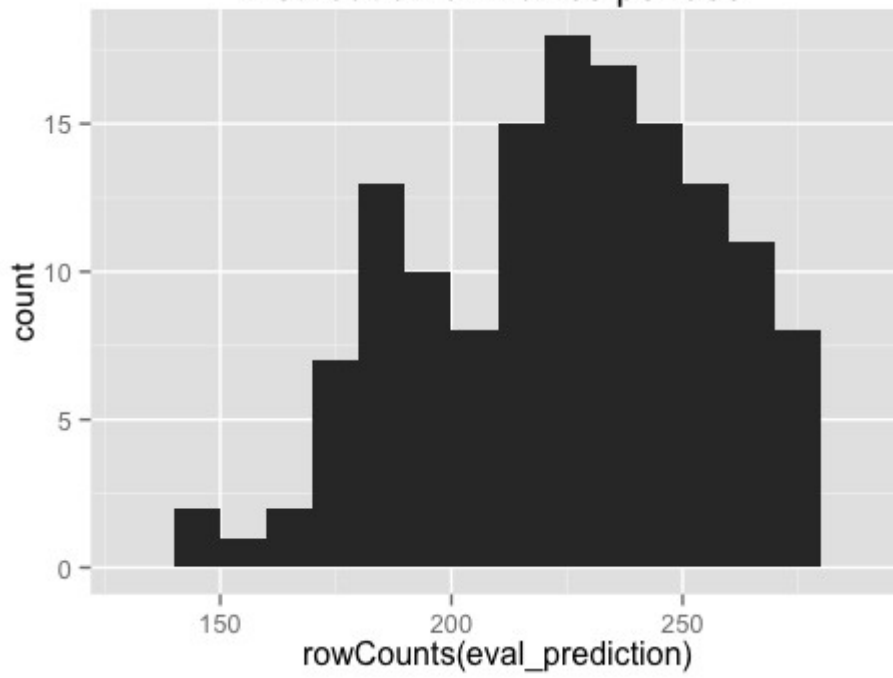
$$distance(item_1, item_2) = \frac{item_1 \cap item_2}{item_1 \cup item_2}$$

$$distance(user_1, user_2) = \frac{user_1 \cap user_2}{user_1 \cup user_2}$$

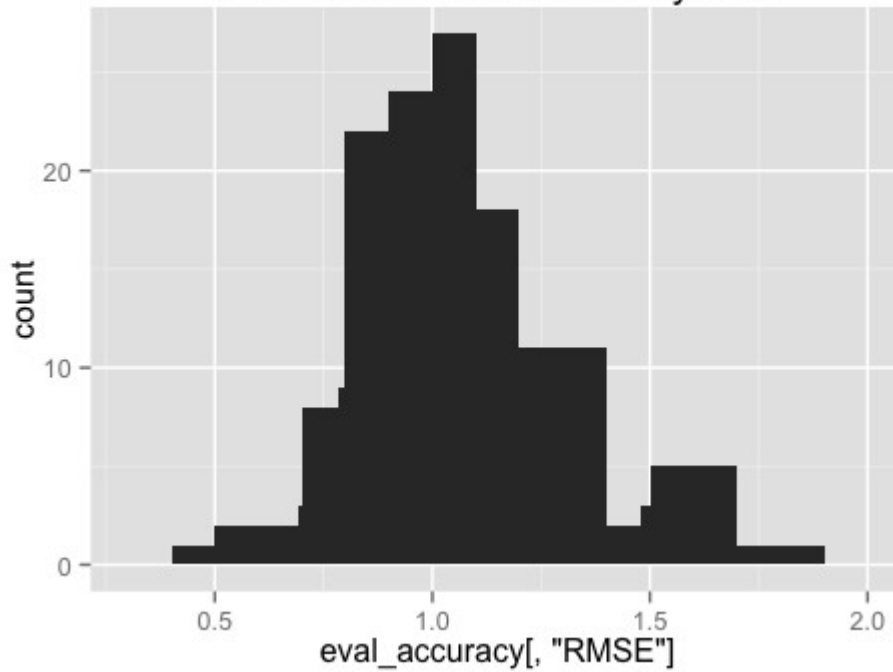
Chapter 4, Evaluating the Recommender Systems



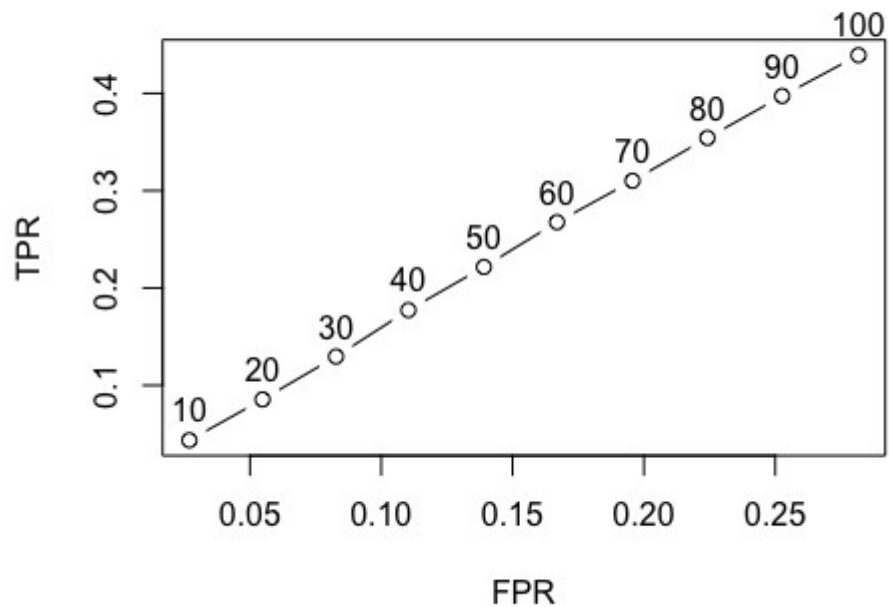
Distribution of movies per user



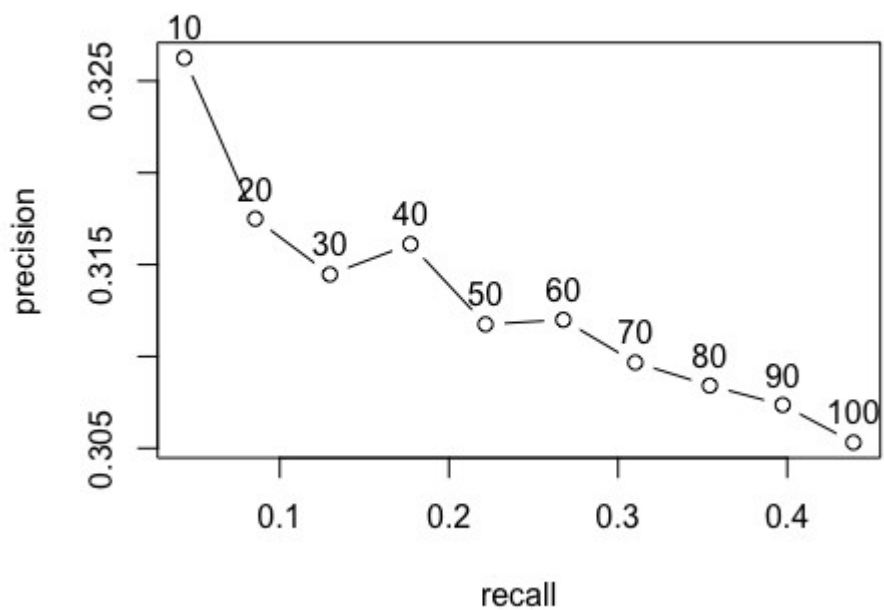
Distribution of the RMSE by user



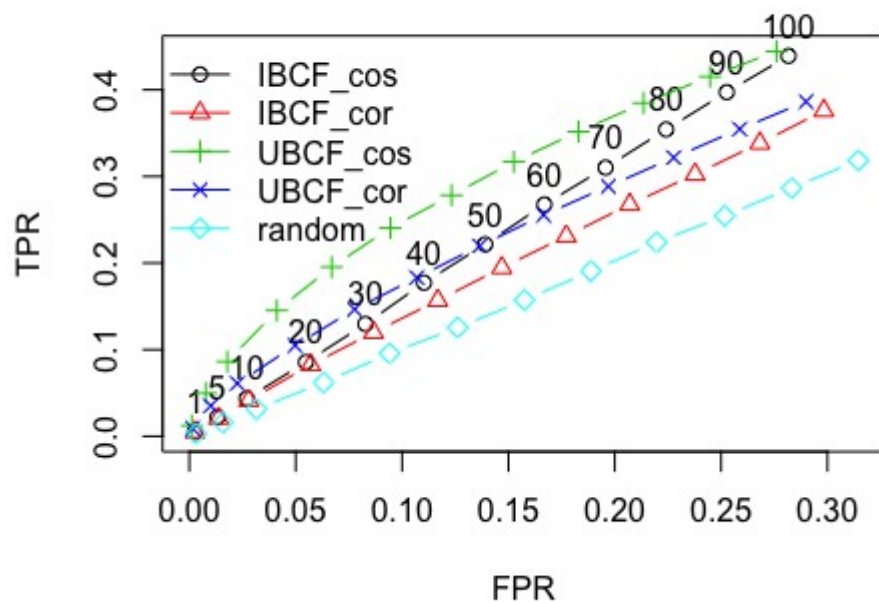
ROC curve



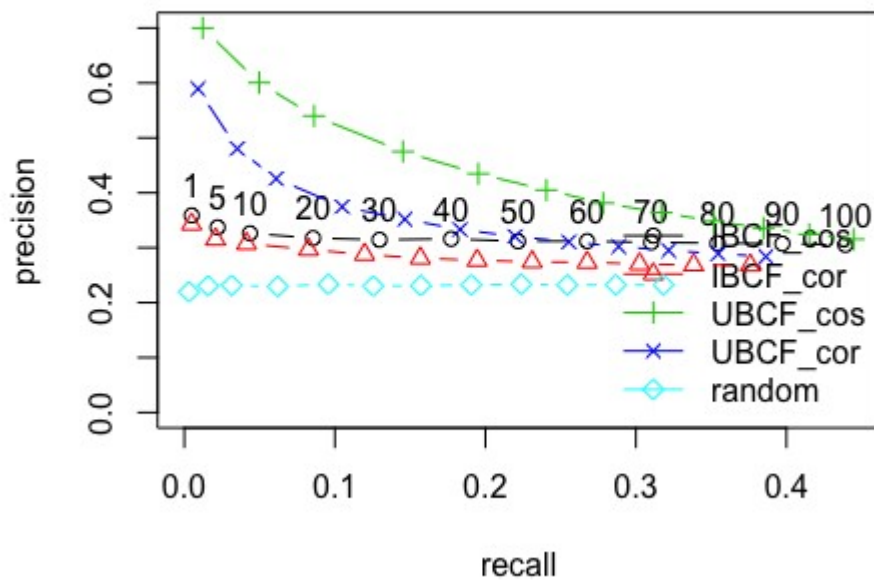
Precision-recall



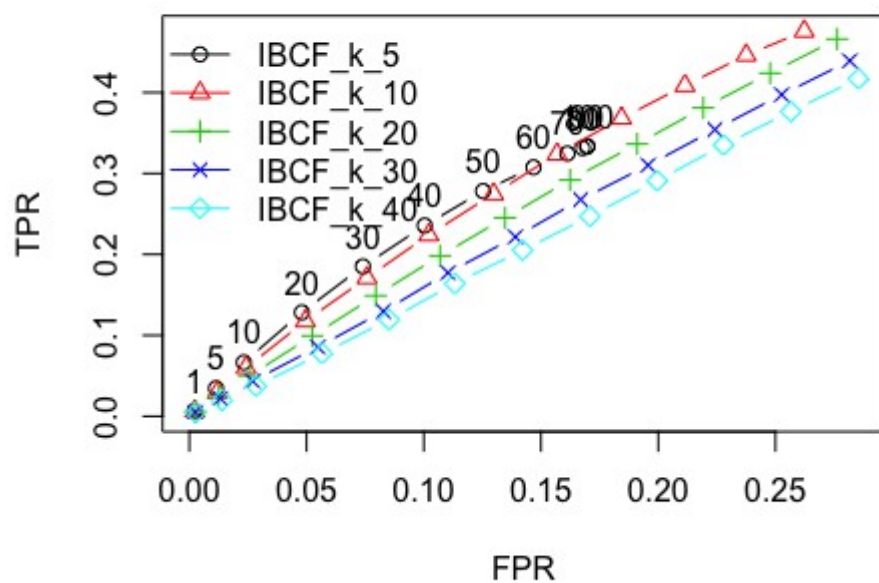
ROC curve



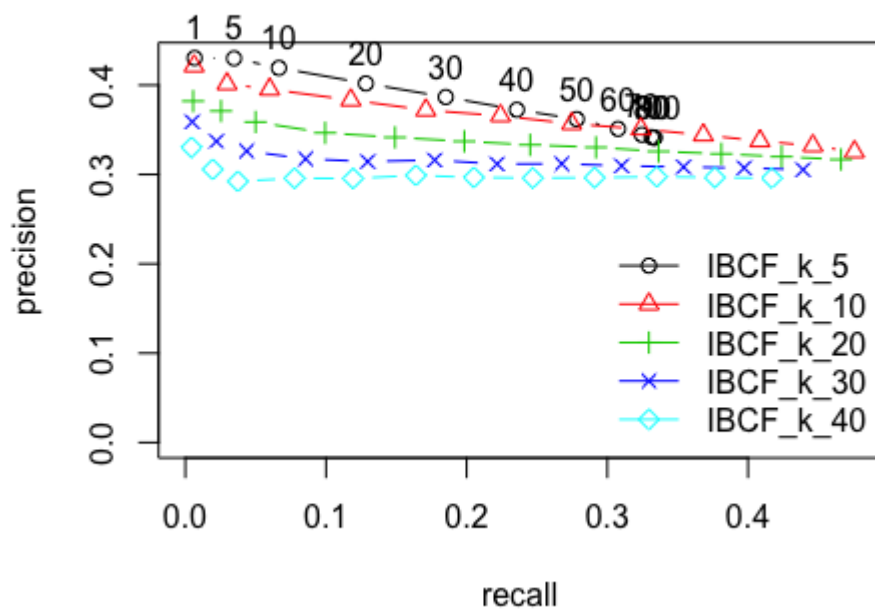
Precision-recall



ROC curve

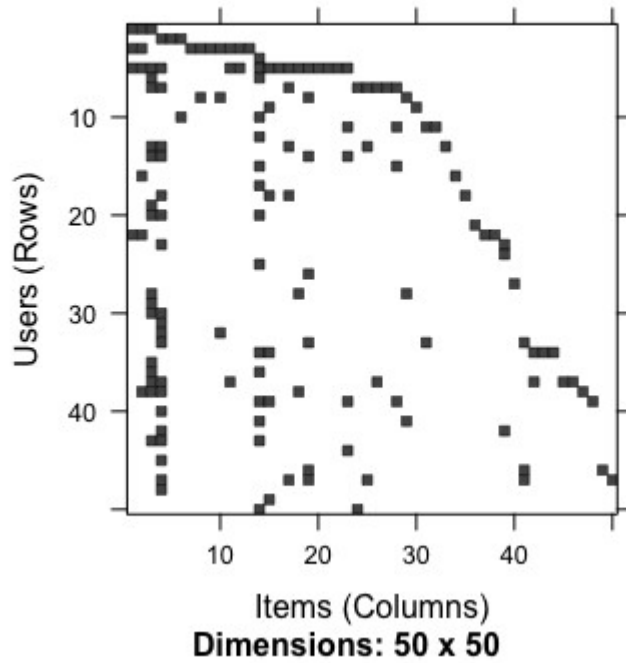


Precision-recall



Chapter 5, Case Study - Building Your Own Recommendation Engine

Binary rating matrix



Distribution of the number of users

