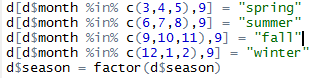
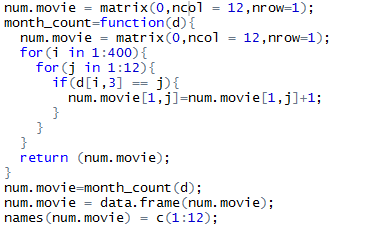
데이터 가공

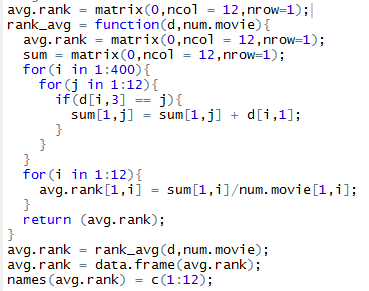


Csv 파일을 읽어 d에 저장한다.

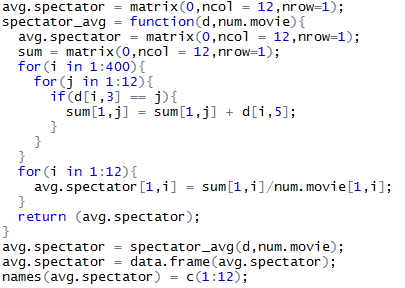
d안의 company부분은 factor형으로 바꿔주고 season이라는 열을 하나 추가한다. month값에 따라 d$season의 값을 정하고 factor형으로 바꿔준다.

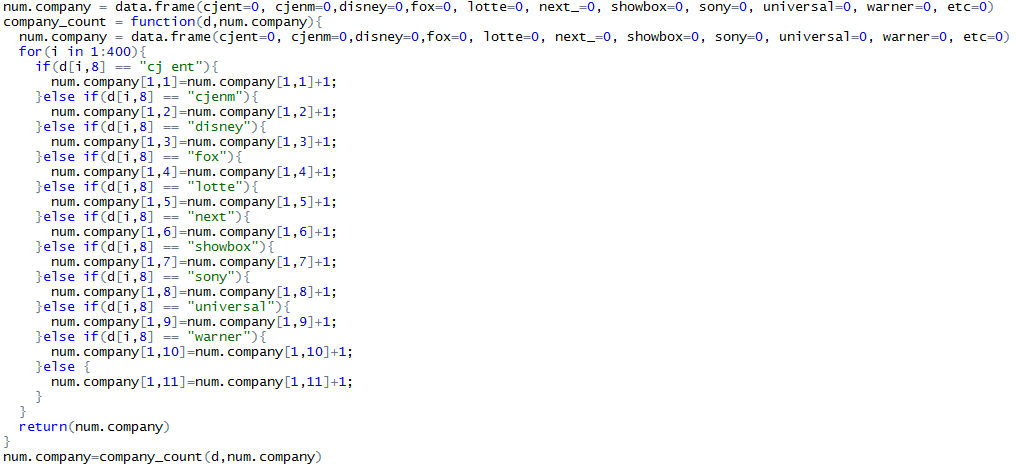


num.movie라는 이름의 행렬을 만들어서 각 월의 개수를 세어서 저장한다. 그 후 data.frame형으로 바꿔준다.

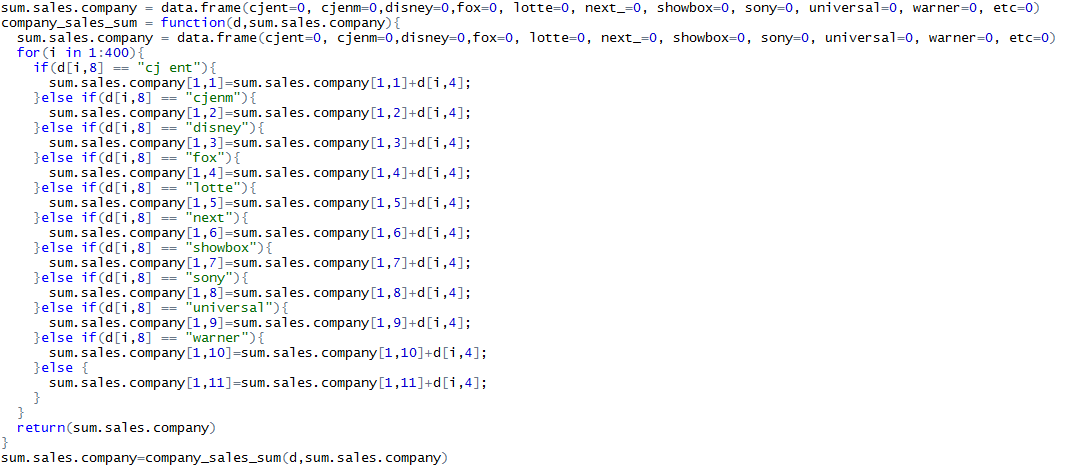


avg.rank라는 이름의 행렬을 만들어서 각 월의 rank의 평균을 계산하여 저장한다. 그 후 data.frame형으로 바꿔준다.

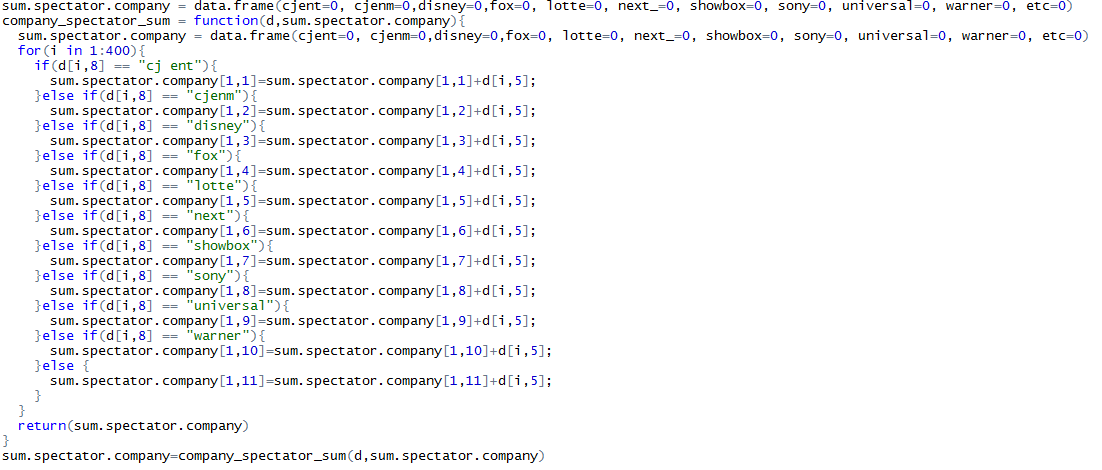


avg.spectator라는 이름의 행렬을 만들어서 각 월의 spectator의 평균을 계산하여 저장한다. 그 후 data.frame형으로 바꿔준다.

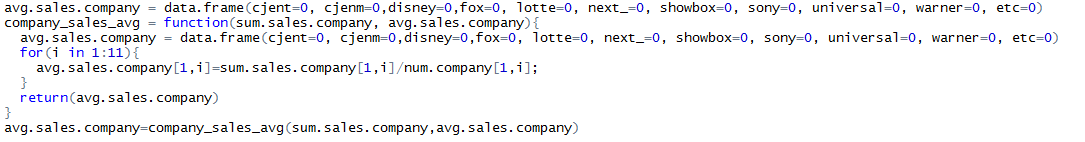
num.company라는 이름의 data.frame에 각 배급사별 영화의 숫자를 계산하여 저장한다.



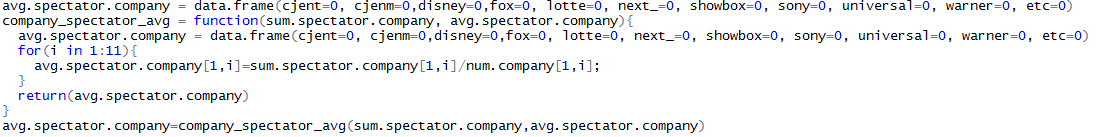
sum.sales.company라는 이름의 data.frame에 각 배급사별 매출액의 총합을 계산하여 저장한다.



sum.spectator.company라는 이름의 data.frame에 각 배급사별 관객의 총합을 계산하여 저장한다.



avg.sales.company라는 이름의 data.frame에 각 배급사별 매출액의 평균을 계산하여 저장한다.



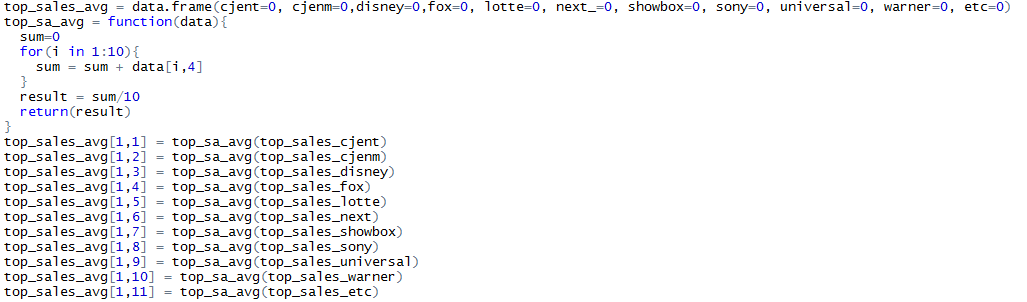
avg.spectator.company라는 이름의 data.frame에 각 배급사별 관객의 평균을 계산하여 저장한다.



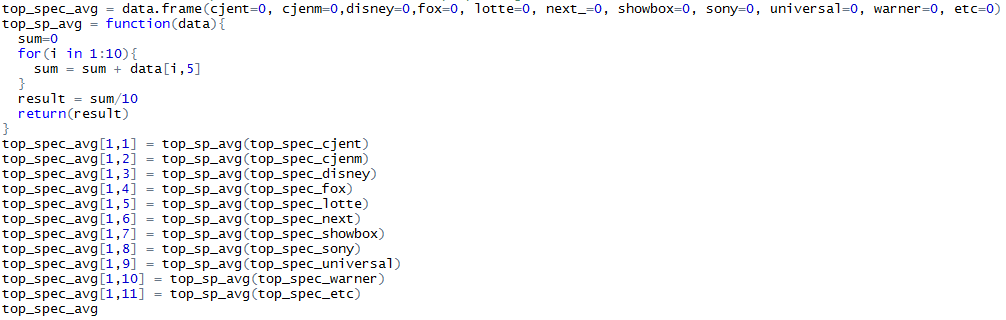
top\_sales\_company명에 각각의 company중 sales가 높은 순으로 10개를 뽑아내 data.frame형태로 저장한다.



top\_spec\_company명에 각각의 company중 spectator가 높은 순으로 10개를 뽑아내 data.frame형태로 저장한다.



각 배급사의 상위 10개의 sales들의 평균을 구하여 top\_sales\_avg라는 data.frame에 저장한다.



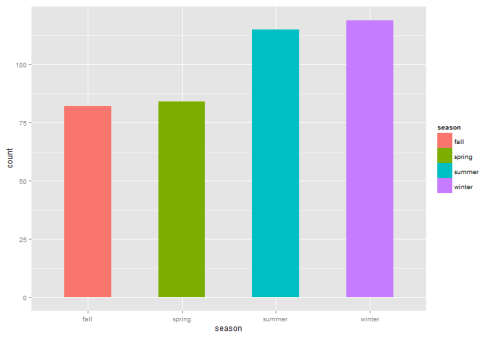
각 배급사의 상위 10개의 spectator들의 평균을 구하여 top\_spec\_avg라는 data.frame에 저장한다.

Visualization



ggplot2를 설치한다.

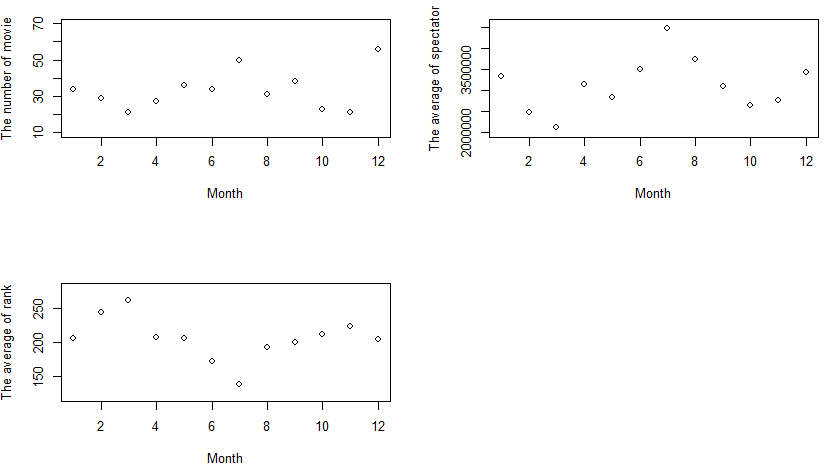


계절별로 나오는 영화의 수를 그래프로 그린다.



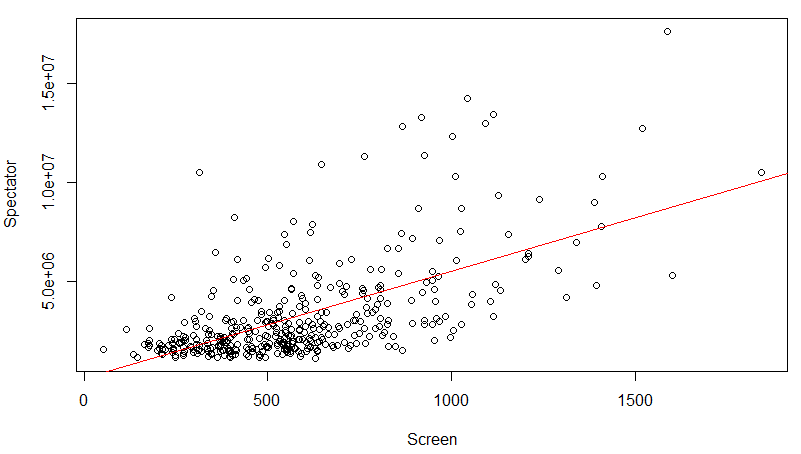


계절별로 영화의 수, 랭크의 평균, 관객수의 평균을 그래프로 그린다.



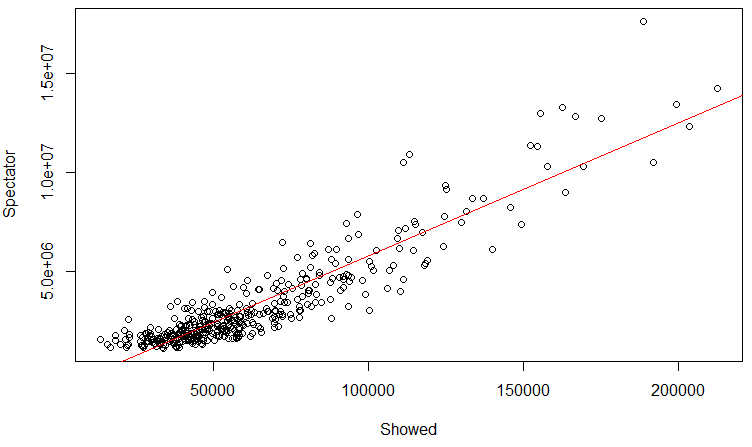


스크린 수와 관객수의 관계를 그리고 추세선을 그린다.



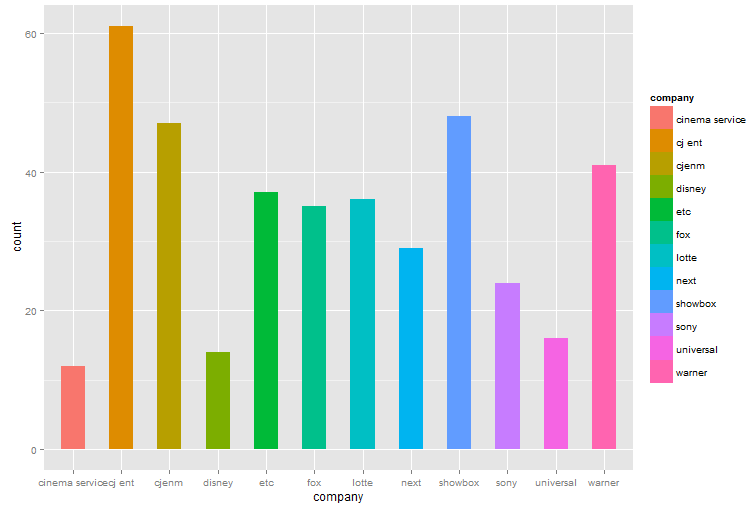


상영횟수와 관객수의 관계를 그리고 추세선을 그린다.



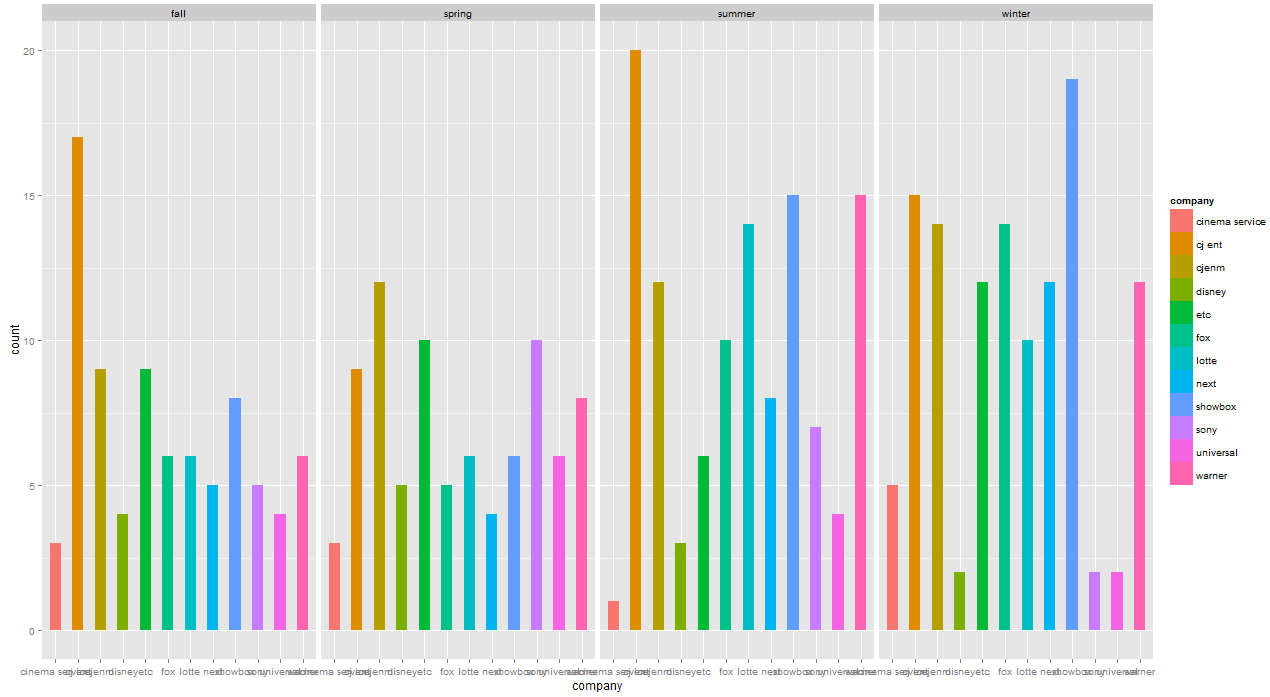


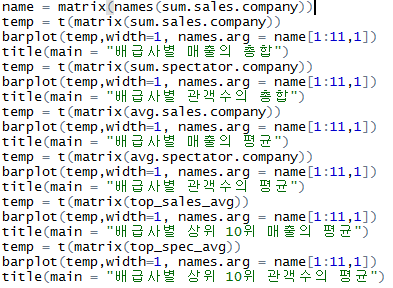
배급사별 출시한 영화의 수를 그린다.





각 계절별 배급사에 따라 출시 된 영화의 수를 그린다.





Barplot을 이용해 배급사별 매출의 총합, 배급사별 관객수의 총합, 배급사별 매출의 평균, 배급사별 관객수의 평균, 배급사별 상위 10위 매출의 평균, 배급사별 상위 10위 관객수의 평균을 그립니다.