Assignment 2

R Markdown

```
#There are three different tables movie, rating and reviewer. movie_id is the primary key in the movie

#Load the information from the three SQL tables into an R dataframe using the select statement.
getresults <- data.frame(dbGetQuery(mydb, "Select movie_name, reviewer_last_name, reviewer_first_name, r
        from movie, reviewer, rating
        where movie.movie_id = rating.movie_id and
        reviewer.reviewer_id = rating.reviewer_id
        order by movie_name, reviewer_last_name, reviewer_first_name, rating"))
```

```
## Warning in .local(conn, statement, ...): Decimal MySQL column 3 imported as
## numeric
```

```
#Display results of the resultant dataframe

getresults
```

```
##             movie_name reviewer_last_name reviewer_first_name rating
## 1        Almost Famous            Caitlin                Andy     NA
## 2        Almost Famous               King         Christopher     NA
## 3        Almost Famous           McDonald               Ellen    4.6
## 4        Almost Famous              Obama              Barack    4.4
## 5        Almost Famous            Rowling                  JK     NA
## 6  Slumdog Millionaire            Caitlin                Andy    4.8
## 7  Slumdog Millionaire               King         Christopher    1.7
## 8  Slumdog Millionaire           McDonald               Ellen     NA
## 9  Slumdog Millionaire              Obama              Barack     NA
## 10 Slumdog Millionaire            Rowling                  JK    3.2
## 11               Tesla            Caitlin                Andy    3.2
## 12               Tesla               King         Christopher     NA
## 13               Tesla           McDonald               Ellen     NA
## 14               Tesla              Obama              Barack     NA
## 15               Tesla            Rowling                  JK    3.1
## 16 The English Patient            Caitlin                Andy    1.7
## 17 The English Patient               King         Christopher    1.5
## 18 The English Patient           McDonald               Ellen     NA
## 19 The English Patient              Obama              Barack     NA
## 20 The English Patient            Rowling                  JK     NA
## 21   The King's Speech            Caitlin                Andy    4.7
## 22   The King's Speech               King         Christopher     NA
## 23   The King's Speech           McDonald               Ellen     NA
## 24   The King's Speech              Obama              Barack    5.0
```

```
## 25   The King's Speech            Rowling                  JK    4.5
## 26             Titanic            Caitlin               Andy     NA
## 27             Titanic               King         Christopher     NA
## 28             Titanic           McDonald               Ellen     NA
## 29             Titanic              Obama              Barack     NA
## 30             Titanic            Rowling                  JK     NA
```

```
#Group results by movie name, this is helpful at first glance to retrieve ratings by reviewer for each
```

```
gb_movie_name <- group_by(getresults, movie_name)
```

```
#But what would be really useful is to see the average rating for each movie...
```

```
avg_movie_review <- summarize(gb_movie_name, count = n(), rating = mean(round(rating), na.rm=TRUE))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
avg_movie_review
```

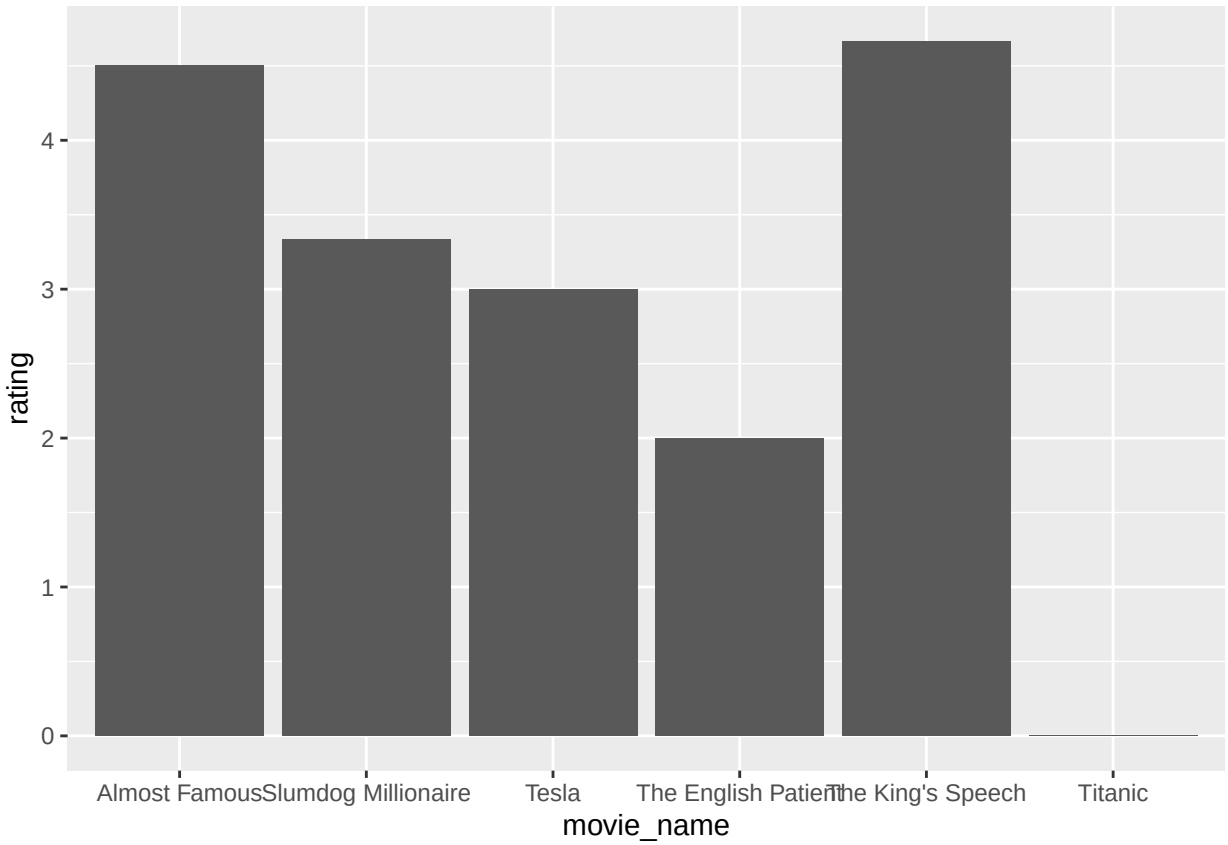
```
## # A tibble: 6 x 3
##   movie_name          count rating
##   <chr>               <int>  <dbl>
## 1 Almost Famous           5   4.5
## 2 Slumdog Millionaire     5   3.33
## 3 Tesla                   5   3
## 4 The English Patient     5   2
## 5 The King's Speech       5   4.67
## 6 Titanic                 5 NaN
```

```
#There is one reviewer who hasn't reviewed any movies and we set that value to 0.
```

```
avg_movie_review$rating[is.nan(avg_movie_review$rating)] <- 0
```

```
#Finally, we plot and analyze the data. Among the survey responses, The King's Speech had the highest r
```

```
ggplot(data = avg_movie_review) + geom_bar(mapping = aes(x = movie_name, y = rating), stat = "identity"
```


rating
4
3
2
1
0
Almost Famous
Slumdog Millionaire
Tesla
The English Patient
The King's Speech
Titanic
movie_name