**INTRODUCTION TO MACHINE LEARNING**

**(NPFL054)**

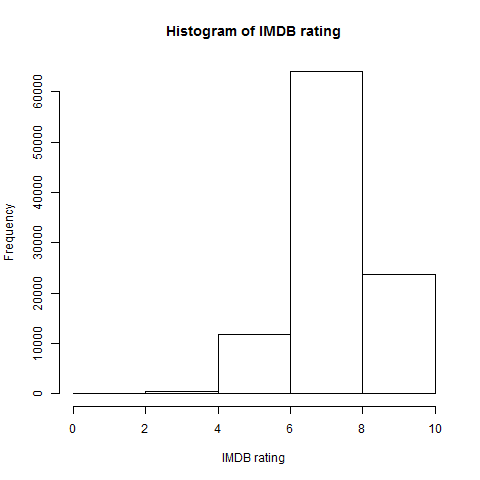
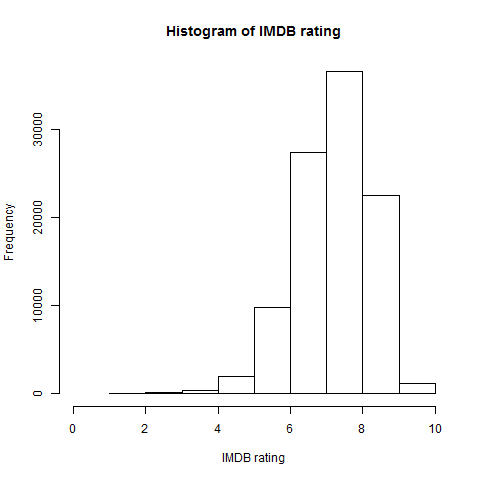
**A template for Homework #1**

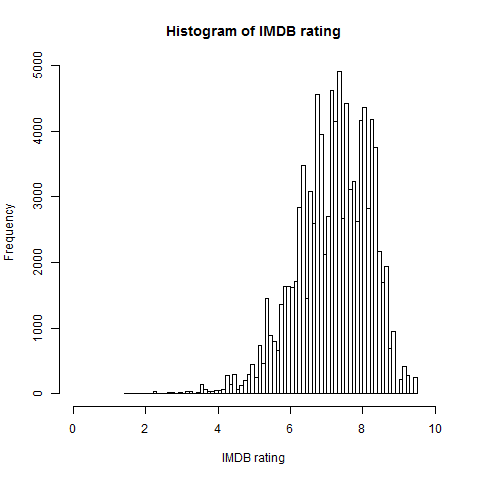
**Name: Lukáš Jendele**

**School year: 2016/17**

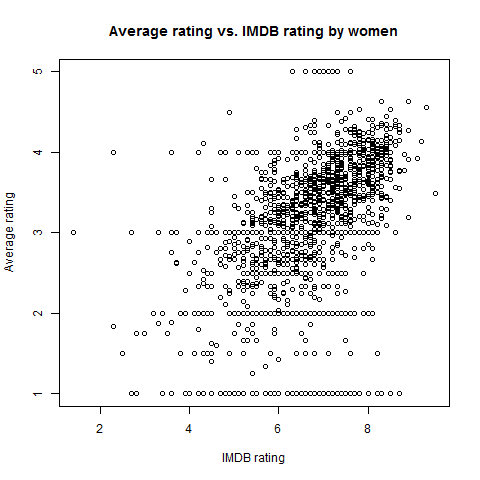
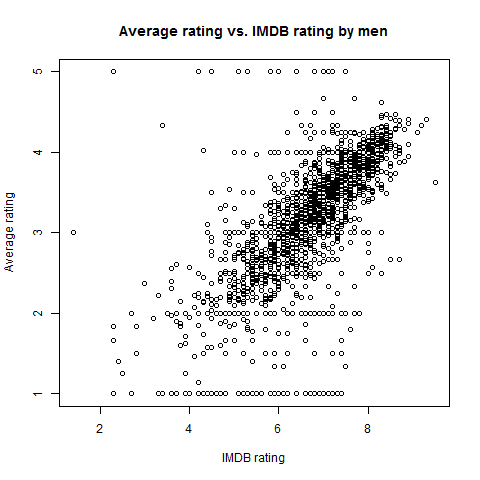
* **Provide answers for the exercises (1) - (6).**
* **For each exercise, your answer cannot exceed one sheet of paper.**

**1. Histograms of imdb\_rating**

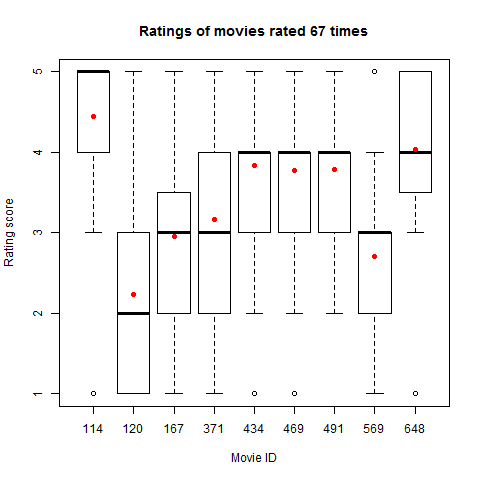
****



**2. Scatterplot matrix of the IMDb rating and average rating**



**3. Boxplots of ratings of the movies rated 67 times**



**4. Voting periods**

Program output:

Interval: 1 Duration: 18.68 days Number of movies rated: 1123

Interval: 2 Duration: 30.38 days Number of movies rated: 1206

Interval: 3 Duration: 9.53 days Number of movies rated: 1130

Interval: 4 Duration: 12.25 days Number of movies rated: 1120

Interval: 5 Duration: 22.94 days Number of movies rated: 1210

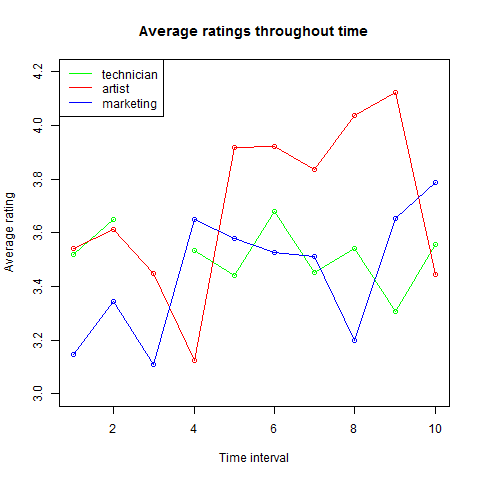
Interval: 6 Duration: 21.38 days Number of movies rated: 1192

Interval: 7 Duration: 27.38 days Number of movies rated: 1337

Interval: 8 Duration: 25.44 days Number of movies rated: 1303

Interval: 9 Duration: 24.83 days Number of movies rated: 1200

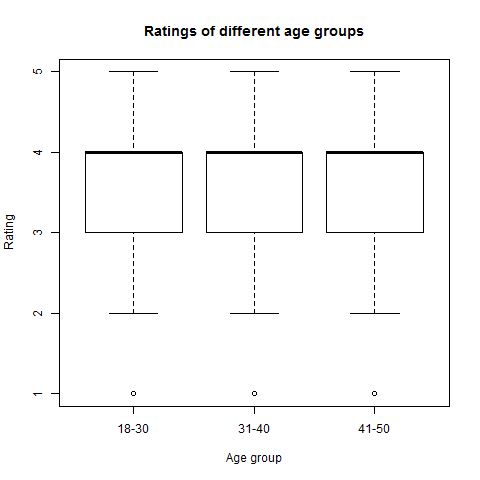
Interval: 10 Duration: 22.04 days Number of movies rated: 1343



**5. Age groups**

|  |  |  |
| --- | --- | --- |
| Age group: | Most frequent occupation | Number of votes |
| 18-30 | Student | 18531 |
| 31-40 | Educator | 3331 |
| 41-50 | Educator | 2896 |

Interpretation of the boxplot below:



*Feature age and target value (rating) are independent.*

**6. Feature frequency of genre\_\* features**

|  |  |
| --- | --- |
| Genre | Frequency |
| Action | 25589 |
| Adventure | 13753 |
| Animation | 3605 |
| Children | 7182 |
| Comedy | 29832 |
| Crime | 8055 |
| Documentary | 758 |
| Drama | 39895 |
| Fantasy | 1352 |
| Filmnoir | 1733 |
| Horror | 5317 |
| Musical | 4954 |
| Mystery | 5245 |
| Romance | 19461 |
| Scifi | 12730 |
| Thriller | 21872 |
| War | 9398 |
| Western | 1854 |