Міністерство освіти і науки України

Національний університет “Львівська політехніка”

Інститут комп'ютерних наук та інформаційних технологій

Кафедра систем штучного інтелекту



**Лабораторна робота №2**

*З дисципліни:*

**“** **Видуботок великих даних”**

**Виконала:**

Студентка групи КН-318

Яворська М.О.

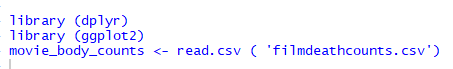
**Перевірила:**

професор Шаховська Н.Б.

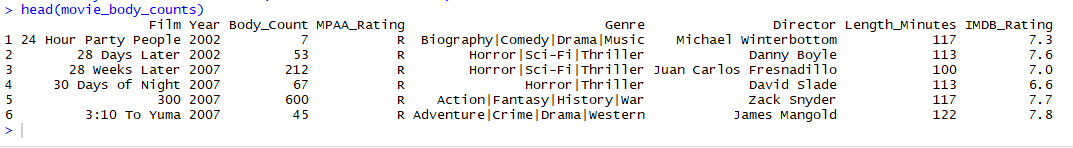
Львів-2019

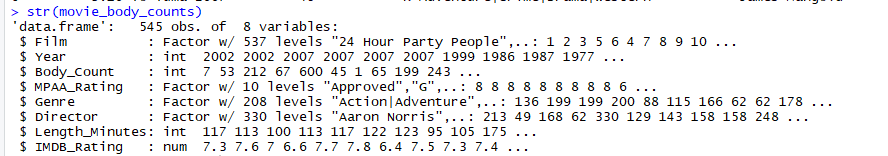
**Хід роботи**

• Load the library and the file:



• Explore the structure of our dataset:



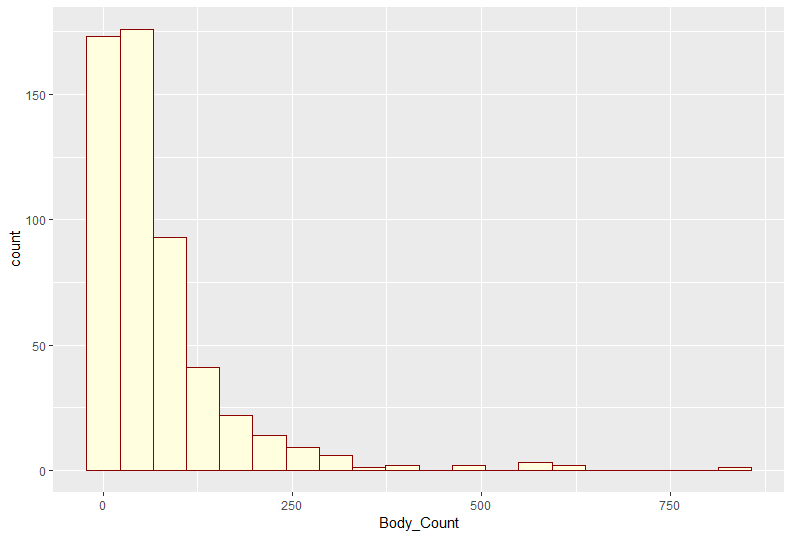


• Add a new field body\_per\_min, which contains the ratio of those killed in the movie to the length of the movie in minutes:

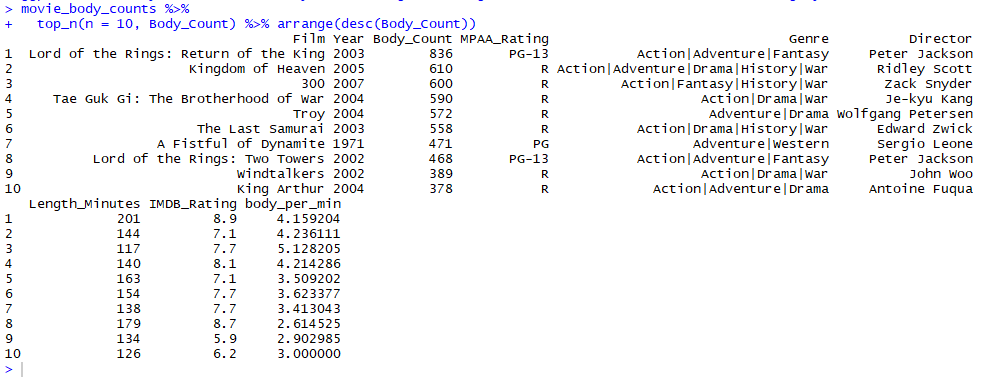


• Construct a histogram for the number of characters who have died:

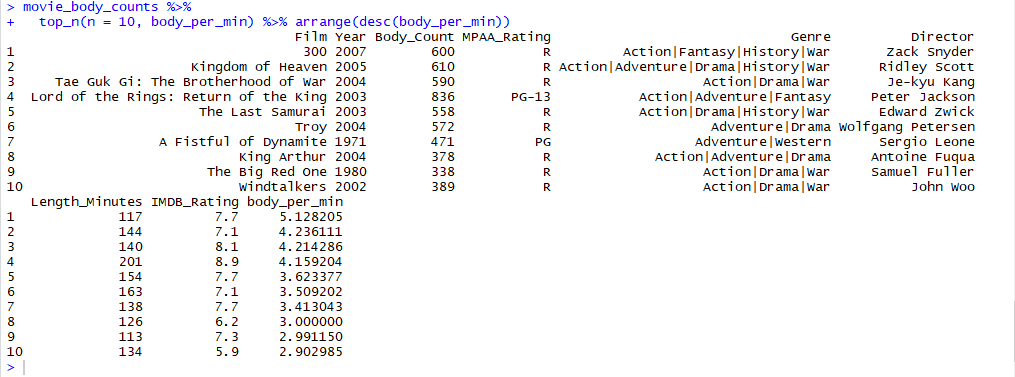




• Find the top 10 films, killing most of the characters:

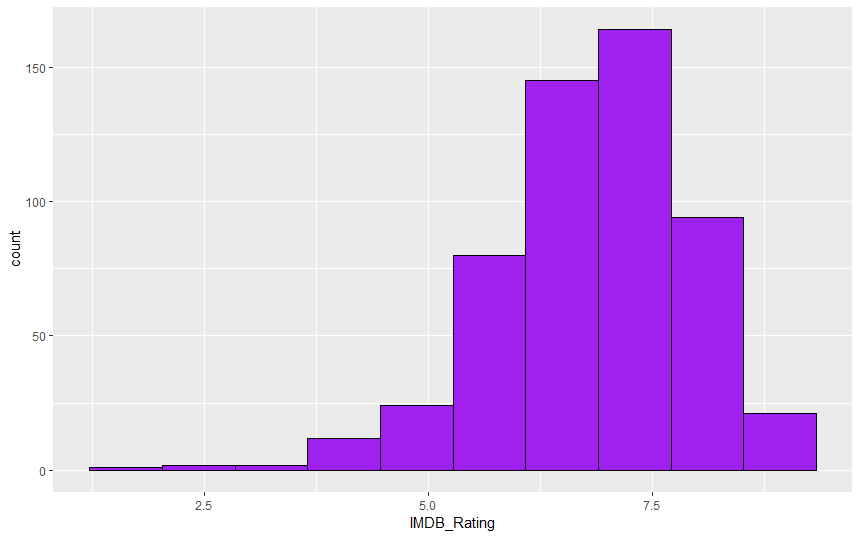


• And movies, killing the largest number of characters in relation to the length of the film:

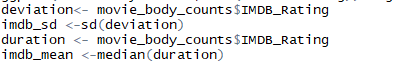


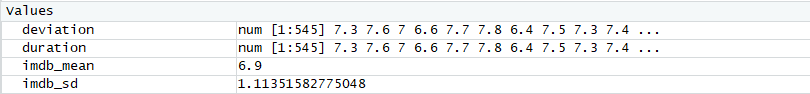
• Construct a histogram IMDB rating:





• Find the mean and standard deviation for the variable IMDBrating, variables and give names





• Let's generate a normal distribution that has a mean value and standard deviation imdb\_mean imdb\_sd. To use this feature rnorm. To sequence was generated constant at each performance of our code, set set.seed

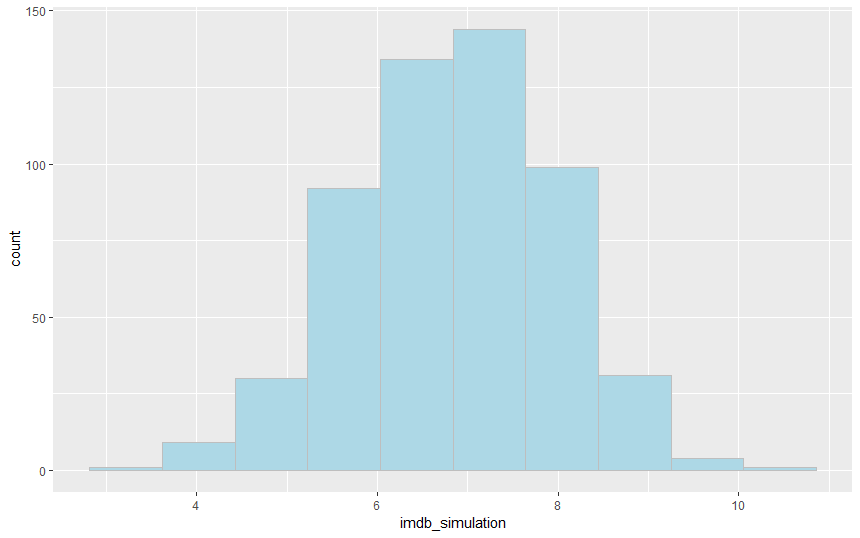
 

• Add these values to our table:



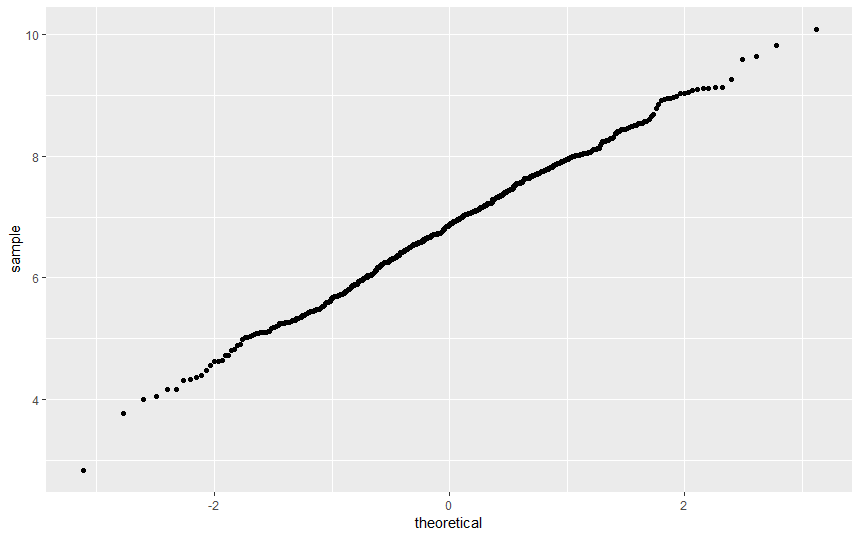


• Construct a histogram for the simulation:



• To test whether the normal distribution, use function qqplot. Let's use it to check whether the normally distributed data rating IMDB. First qqplot build our simulation





• And now for the real rating IMDB\_Rating:



