Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**(ФГБОУ ВО «КубГУ»)**

**Факультет компьютерных технологий и прикладной математики**

**Кафедра вычислительных технологий**

**ЛАБОРАТОРНАЯ РАБОТА №2**

**Дисциплина: Обработка больших данных**

Работу выполнил: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ К. А. Корнилов

Направление подготовки: 02.03.02 Фундаментальная информатика и информационные технологии

Преподаватель: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_В. И. Шиян

**Тема:** Способы подготовки и отображения данных в R.

**Цель**: Изучить понятие нечеткой классификации предложений.

**Задание:**

1. Выполнить мини соцопрос в своей группе по темам по шкале от 0 до 10 или дробная шкала 0-1, для этого сделать таблицу в Google Docs:
2. По результатам необходимо: вычислить max, min, mean по каждому столбцу, подсчитать количество людей, отдавших предпочтение >0.7 и <0.3(составить вектор),вывести рейтинг фильмов (книг...) в списке по убыванию, построить столбчатую диаграмму оценок (можно сделать разными способами),

**Ход работы**:

1. Был выполнен соц. опрос с использованием Google Forms и Google Docs по теме: “Самый раздражающие вещи в видеоиграх”. Ссылка на опрос: <https://forms.gle/ZyHRigDZT4kANUvU9>.
2. Были поучены в виде Google table и проанализированы результаты опроса. Данные были загружены в среду RStudio.

Код загрузки данных:

library(openxlsx)

library(magrittr)

ex\_data <- read.xlsx("C:\\Users\\kiril\\Desktop\\Rdata\\lab2.xlsx", 1)

colnames(ex\_data)

colnames(ex\_data) = c( "ФИО", "Bugs", "Hitboxes", "AI", "Optimization", "Interface", "Allies", "Donate", "Cutscenes", "Random", "Levelling", "DLC", "Openworld" )

Результат загрузки данных представлен на рисунке 1:

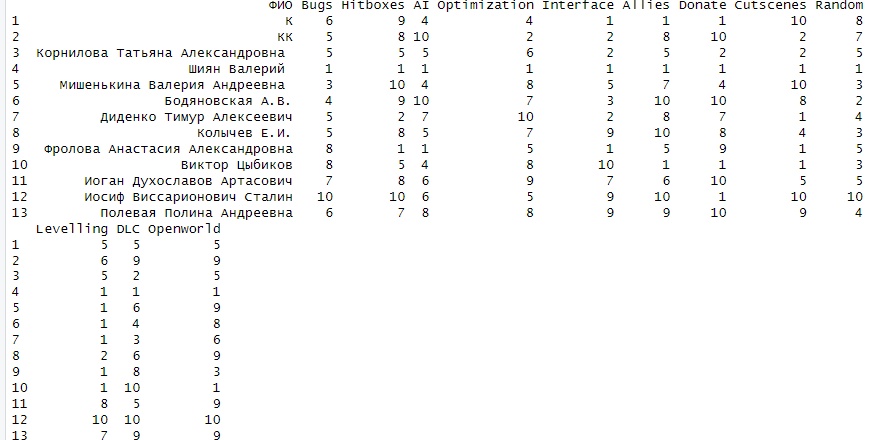


Рисунок 1 – Результат выгрузки данных.

* 1. Были подсчитаны значения max,min,mean по каждому из столбцов.

Код для подсчета:

apply(ex\_data[,c(2:length(colnames(ex\_data)))],2,min)

apply(ex\_data[,c(2:length(colnames(ex\_data)))],2,max)

apply(ex\_data[,c(2:length(colnames(ex\_data)))],2,mean)

Результат подсчета представлен на рисунке 2:

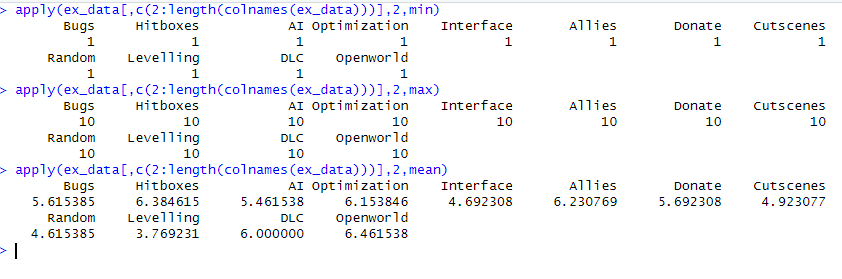


Рисунок 2 – Результат подсчета значений.

* 1. Было подсчитано количество людей по каждому столбцу, отдавших предпочтение >7 и <3 и составлен вектор с соответствующими количествами людей.

Код для подсчета:

as.vector(apply(apply(ex\_data[,c(2:length(colnames(ex\_data)))],2,function(x) x<3),2,sum))

as.vector(apply(apply(ex\_data[,c(2:length(colnames(ex\_data)))],2,function(x) x>7),2,sum))

Результат подсчета представлен на рисунке 3:

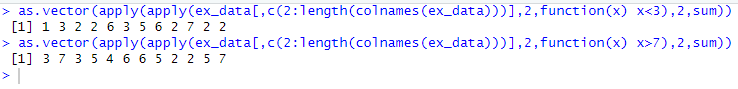


Рисунок 3 – Результат подсчета значений.

* 1. Был выведен рейтинг вещей в списке по убыванию.

Код для вывода рейтинга:

apply(ex\_data[,c(2:length(colnames(ex\_data)))],2,mean) %>% sort (decreasing = TRUE)

Результат подсчета представлен на рисунке 4:

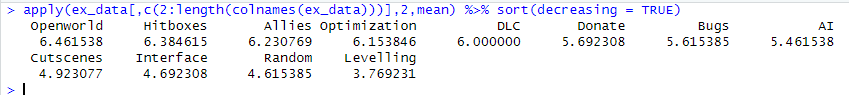


Рисунок 4 – Результат подсчета значений.

* 1. Была построена столбчатая диаграмма оценок.

Код построения столбчатой диаграммы:

barplot(apply(ex\_data[,c(2:length(colnames(ex\_data)))], 2, mean),xlab="Объекты оценки",cex.names=0.6,ylab="Суммарная оценка")

Результат построения представлен на рисунке 5:

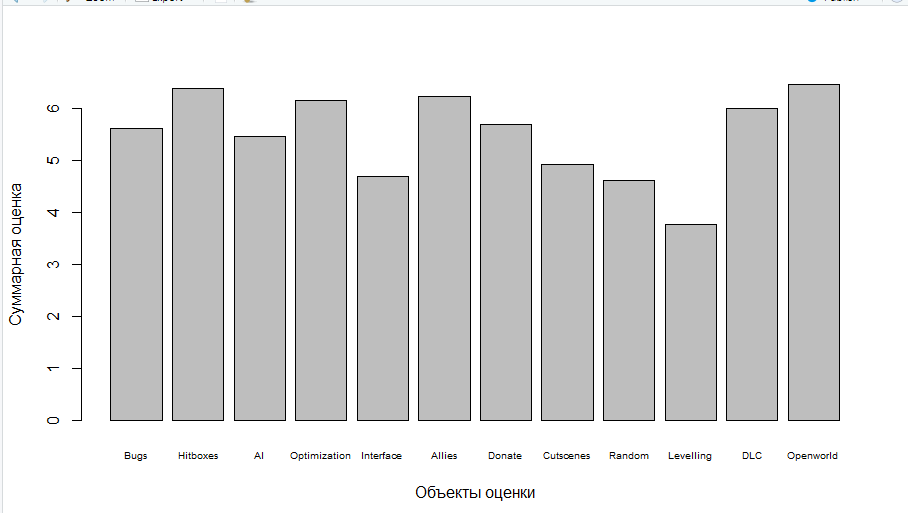
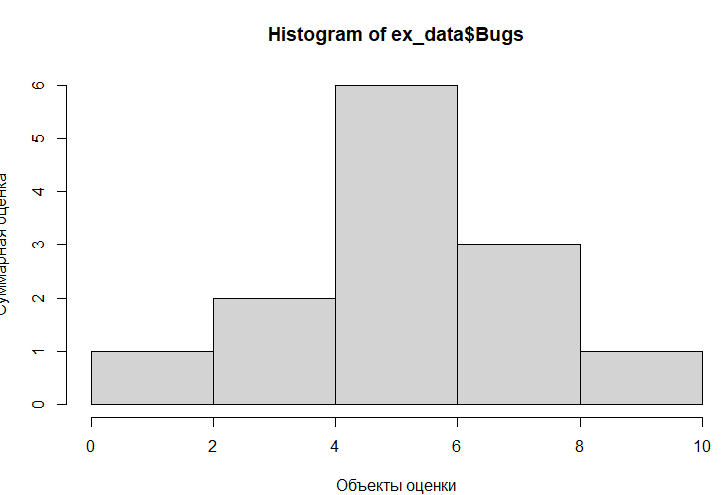
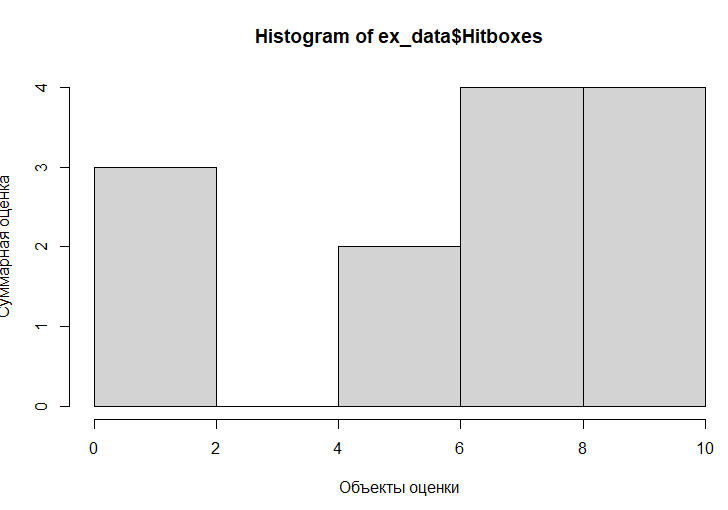
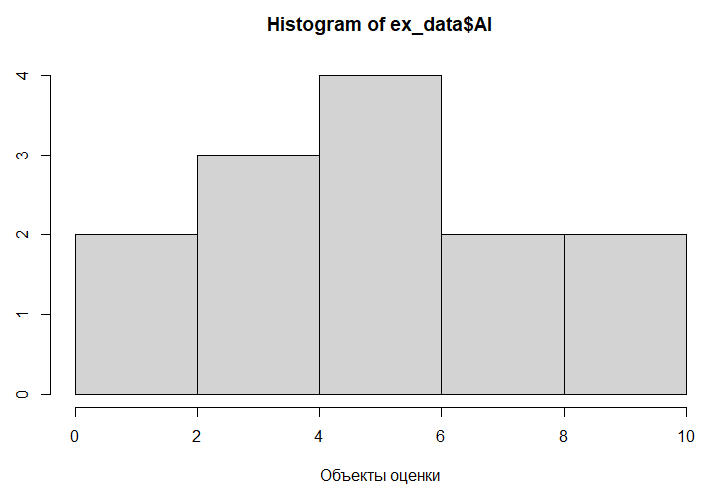
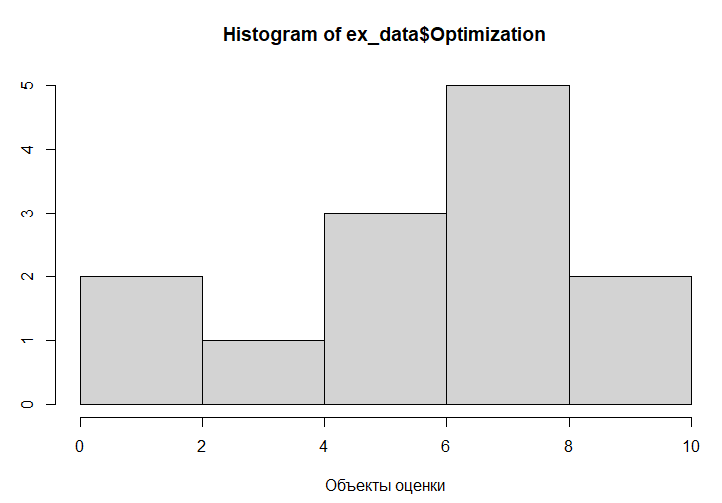


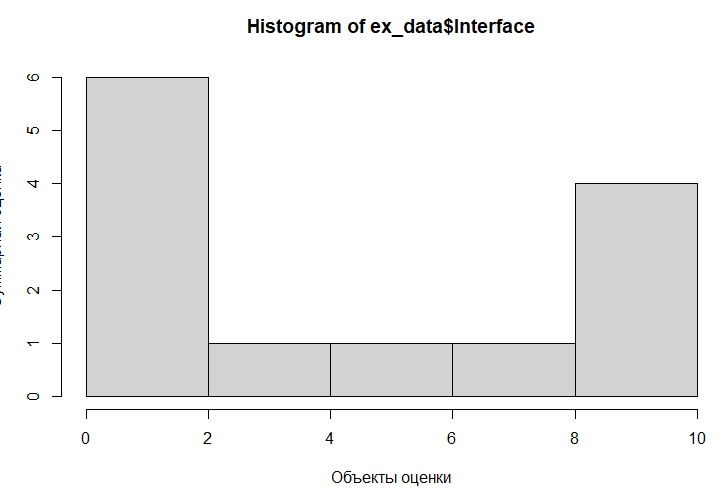
Рисунок 5 – Результат построения диаграммы оценок.

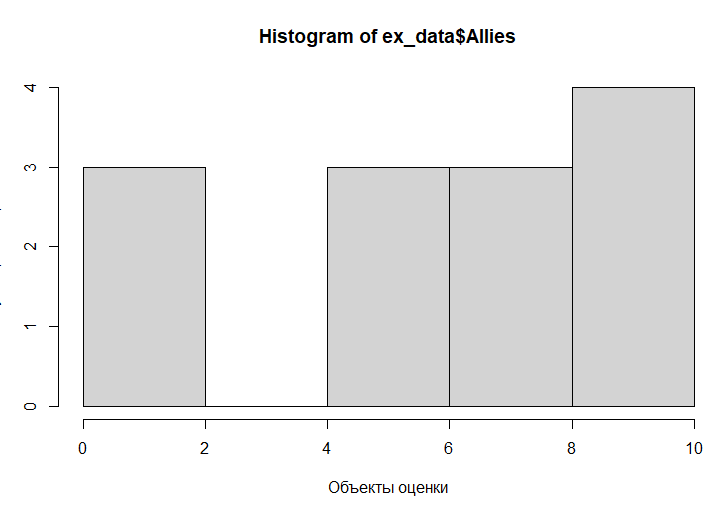


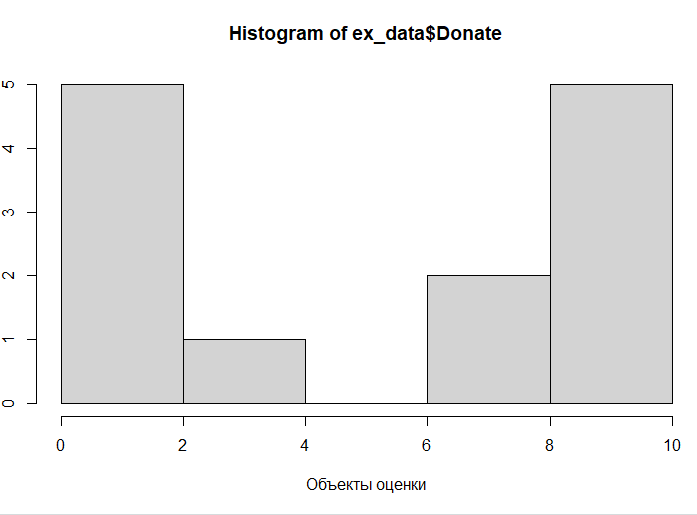


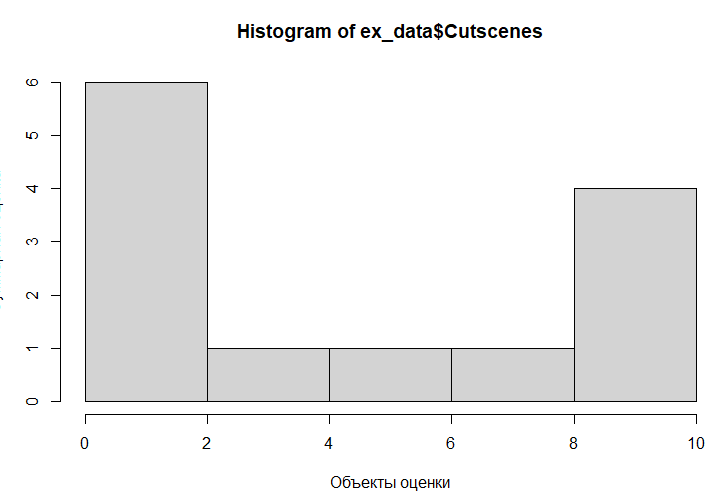


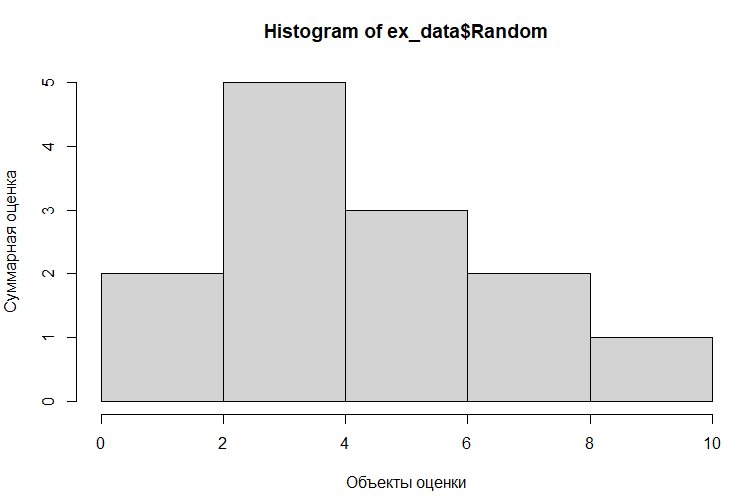


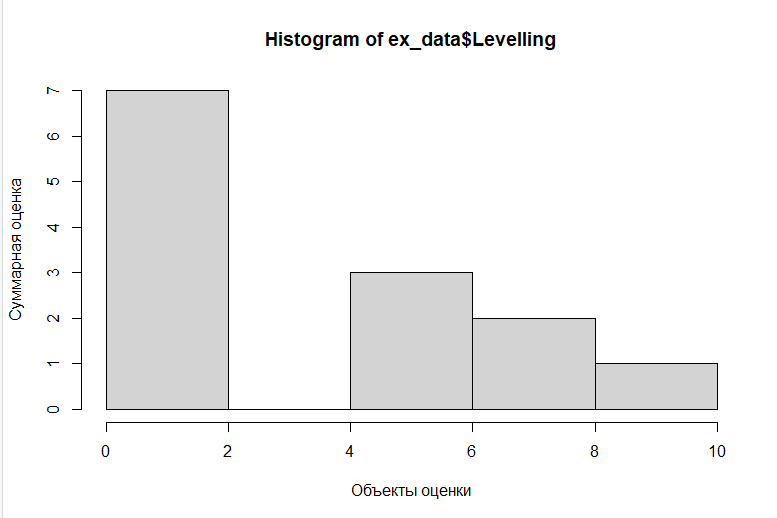


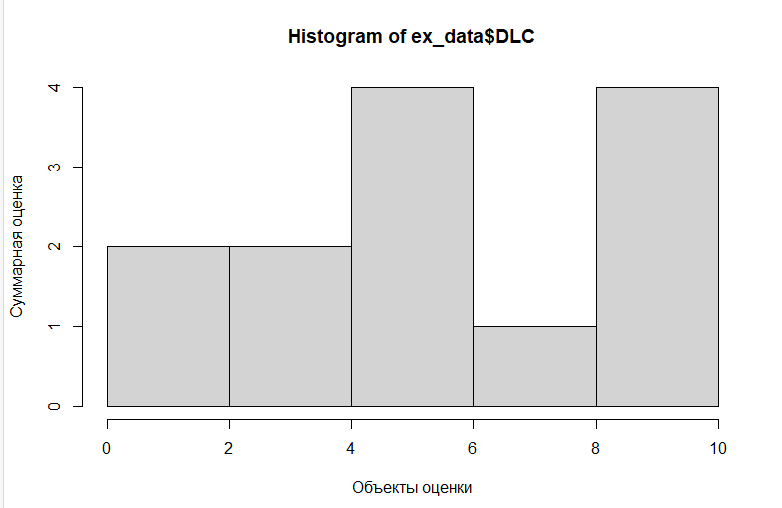












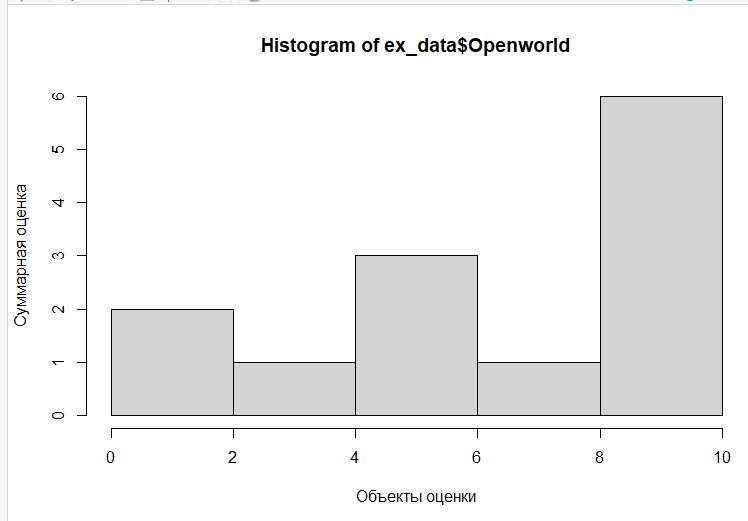


Рисунок 5,6,7,8,9,10,11,12,13,14,15,16 – Диаграмма распределения оценок для каждого из столбцов.

**Вывод:** Были получены начальные навыки анализа данных. С использованием данных навыков был проведен анализ данных на тему “Самый раздражающие вещи в видеоиграх”.