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**

**Отчёт лабораторная №2.2**

**Дисциплина: ОБРАБОТКА БОЛЬШИХ ДАННЫХ**

Работу выполнил \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Кличенко Д. А.

Направление подготовки 02.03.02 Фундаментальная информатика и

информационные технологии

Преподаватель \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Яхонтов А. А.

**Цель работы:** Научиться получать элементарные характеристики наборов данных. Усвоить способы импорта данных. Научиться трактовать некоторые элементарные способы графического представления наборов данных. Выполнить дескриптивный анализ данных.

Тема: «Любимый энергетик»

****

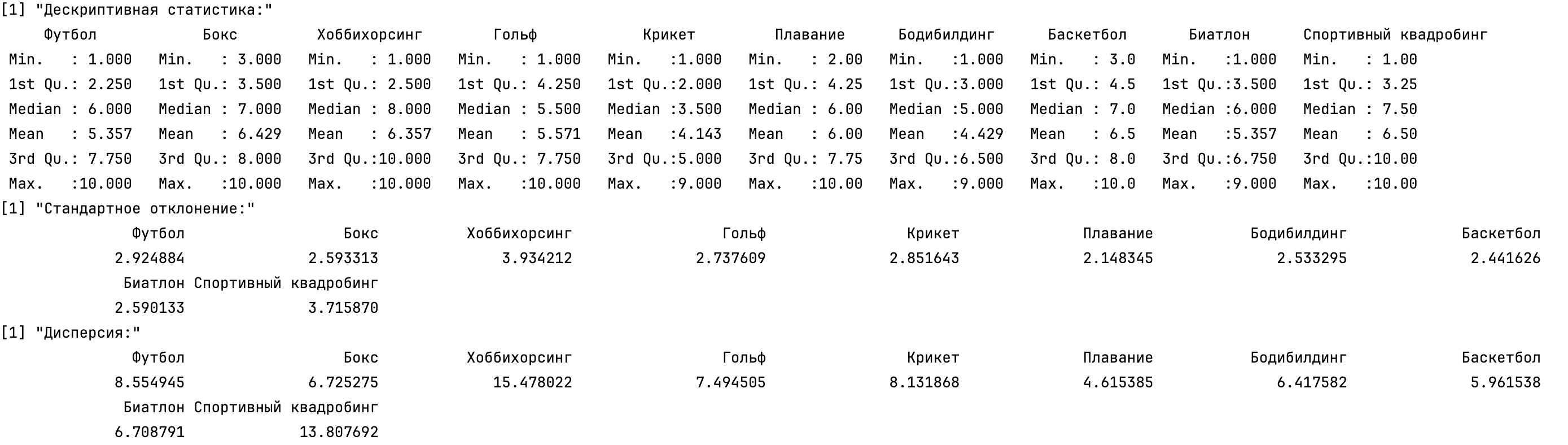
**Задания**

1. Выполнить учебный импорт любых таблиц данных из csv-файла и xls-таблицы.



2. Выполнить дескриптивный анализ данных из ЛР No2\_1 (знать, что входит в

дескриптивный анализ).



На рисунке 1 гистограмма отображает распределение оценок, выставленных респондентами хоббихорсингу. Большинство оценок равны 10 баллам, что говорит о высокой популярности данного вида спорта среди опрошенных. Также встречаются оценки 1, 3 и 8, что указывает на разнообразие мнений.

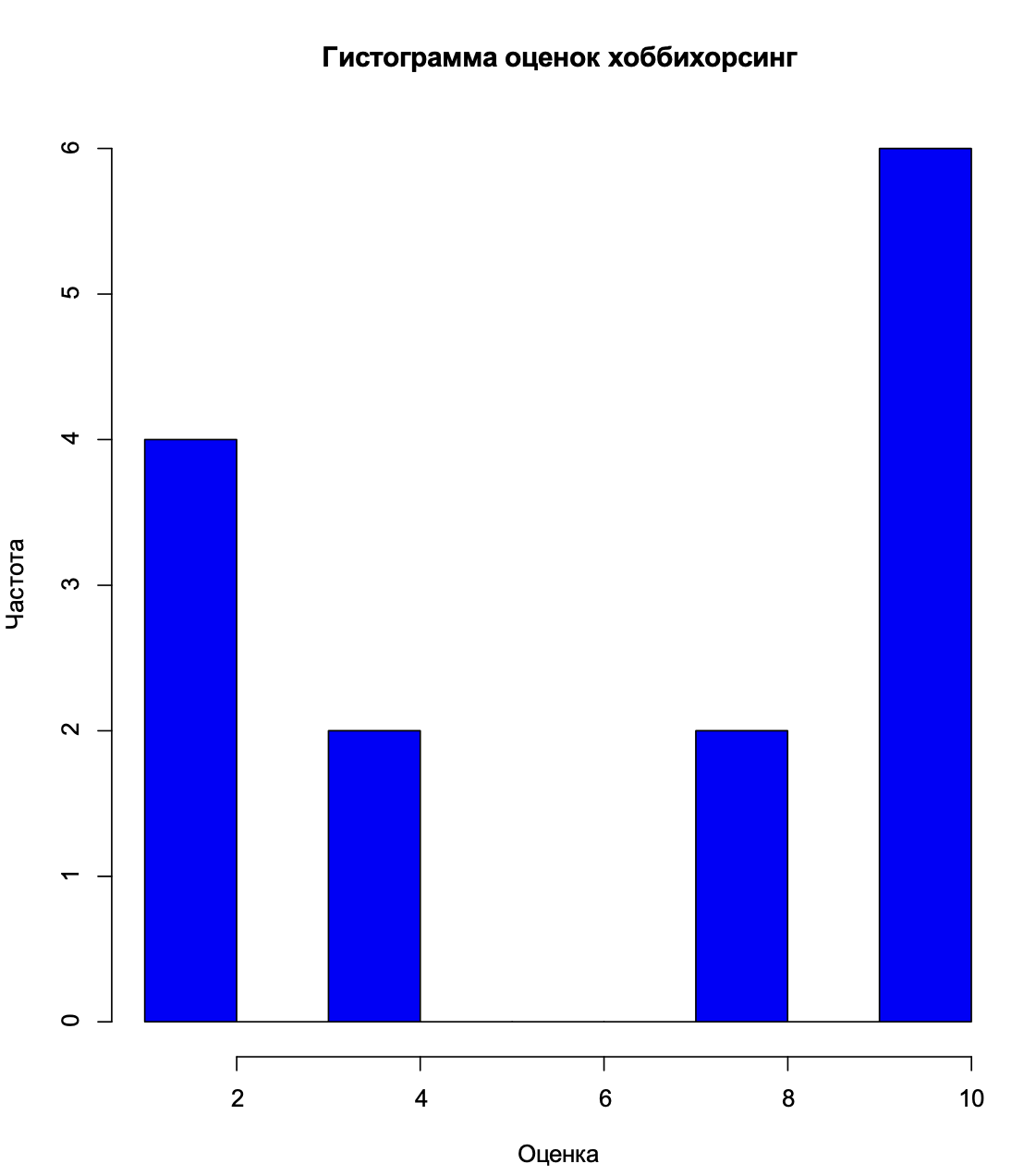


Рисунок 1 – Гистограмма оценок хоббихорсинг

На рисунке 2 изображен боксплот, который показывает распределение оценок разных видов спорта.

* Спортивный квадробинг, Хоббихорсинг,– популярны, медианы высокие.
* Крикет– низкие оценки, маленький разброс (непопулярный).
* Выбросы указывают на редкие, сильно отличающиеся оценки.

Можно сделать вывод, что некоторые виды спорта стабильно получают высокие оценки, другие – вызывают разделённые мнения.

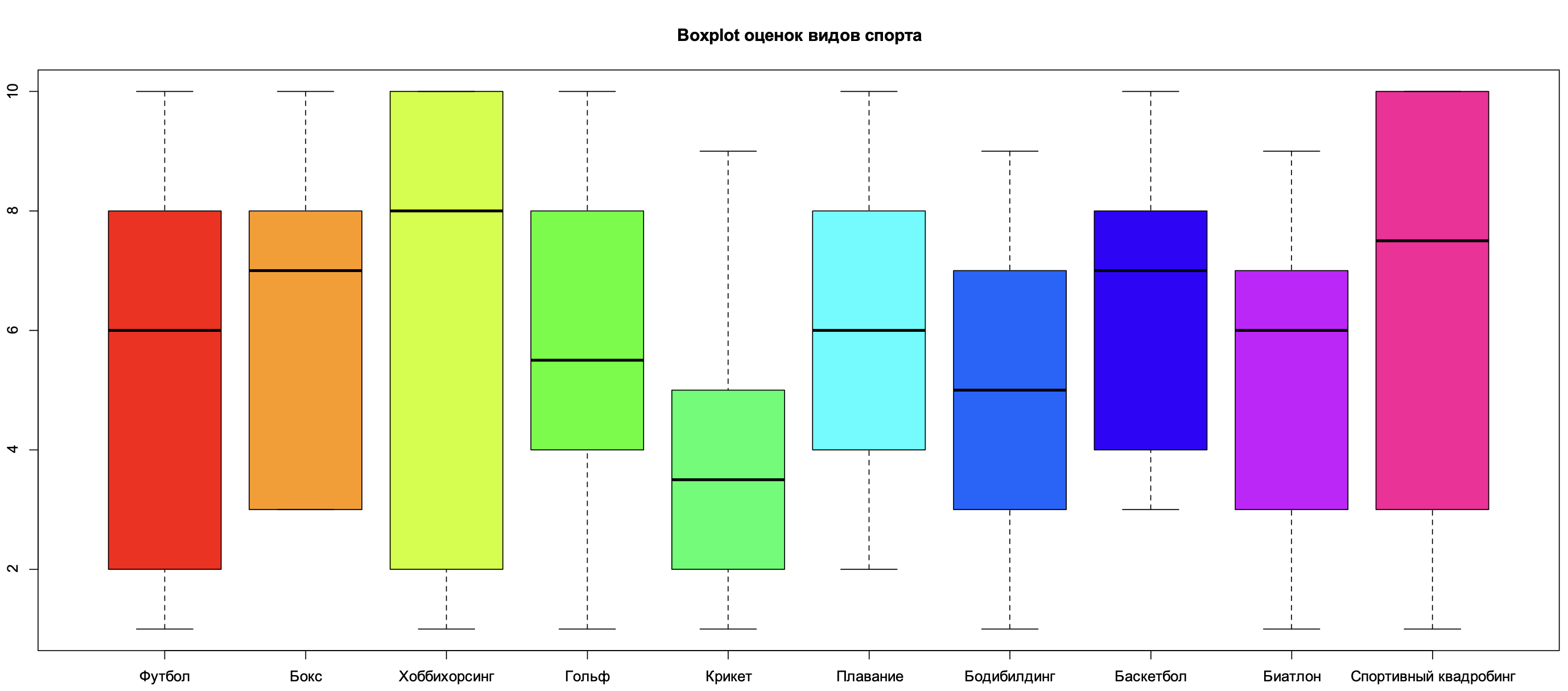
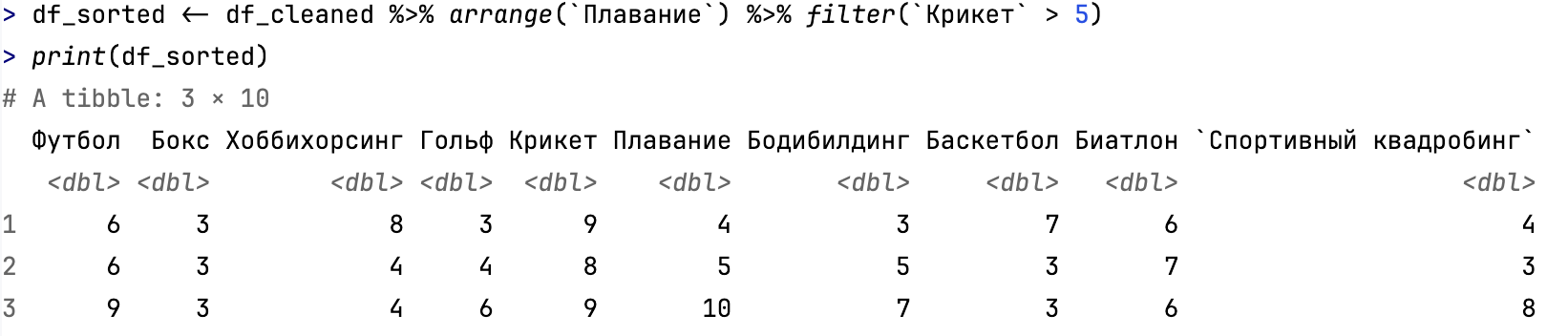
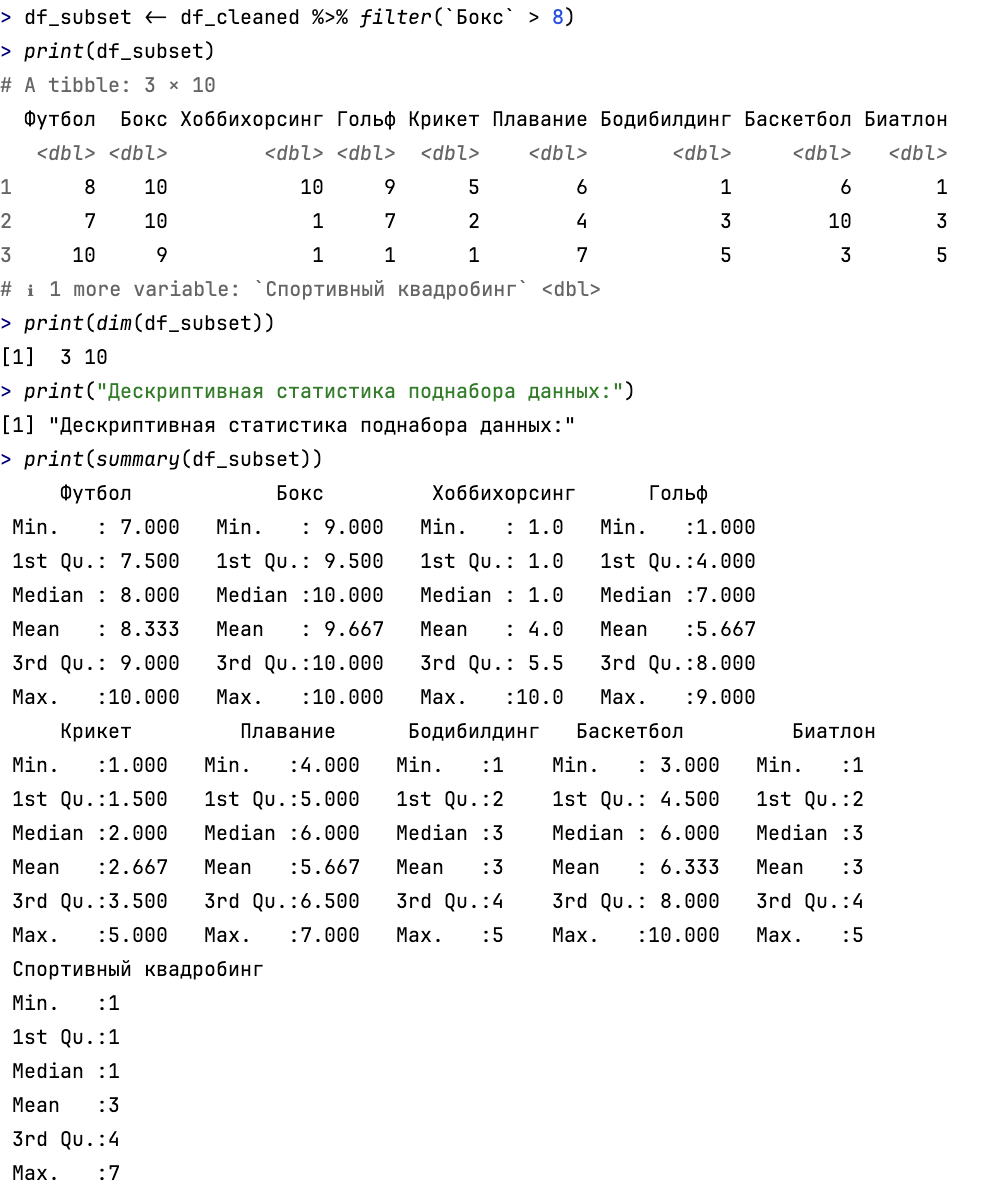


Рисунок 2 - Boxplot оценок видов спорта

3. Выполнить сортировку наборов данных по выбранному признаку.



4. Сформировать отдельные наборы данных по одинаковому признаку (например, составить subdataset, из студентов, отдавших предпочтение по шкале> 0.7 определенной книге), вывести результат, выполнить подсчет размерностей новых таблиц, снова выполнить их анализ – гистограмма, боксплот, серединные меры (см п.2).



На рисунке 3 гистограмма отображает распределение оценок хоббихорсинга в новом наборе данных. Теперь оценки значительно изменились, сохранились только 1 и 10 баллов. После фильтрации данных средние оценки исчезли, мнение респондентов стало полярным: либо данный вид спорта очень нравится, либо совсем нет.

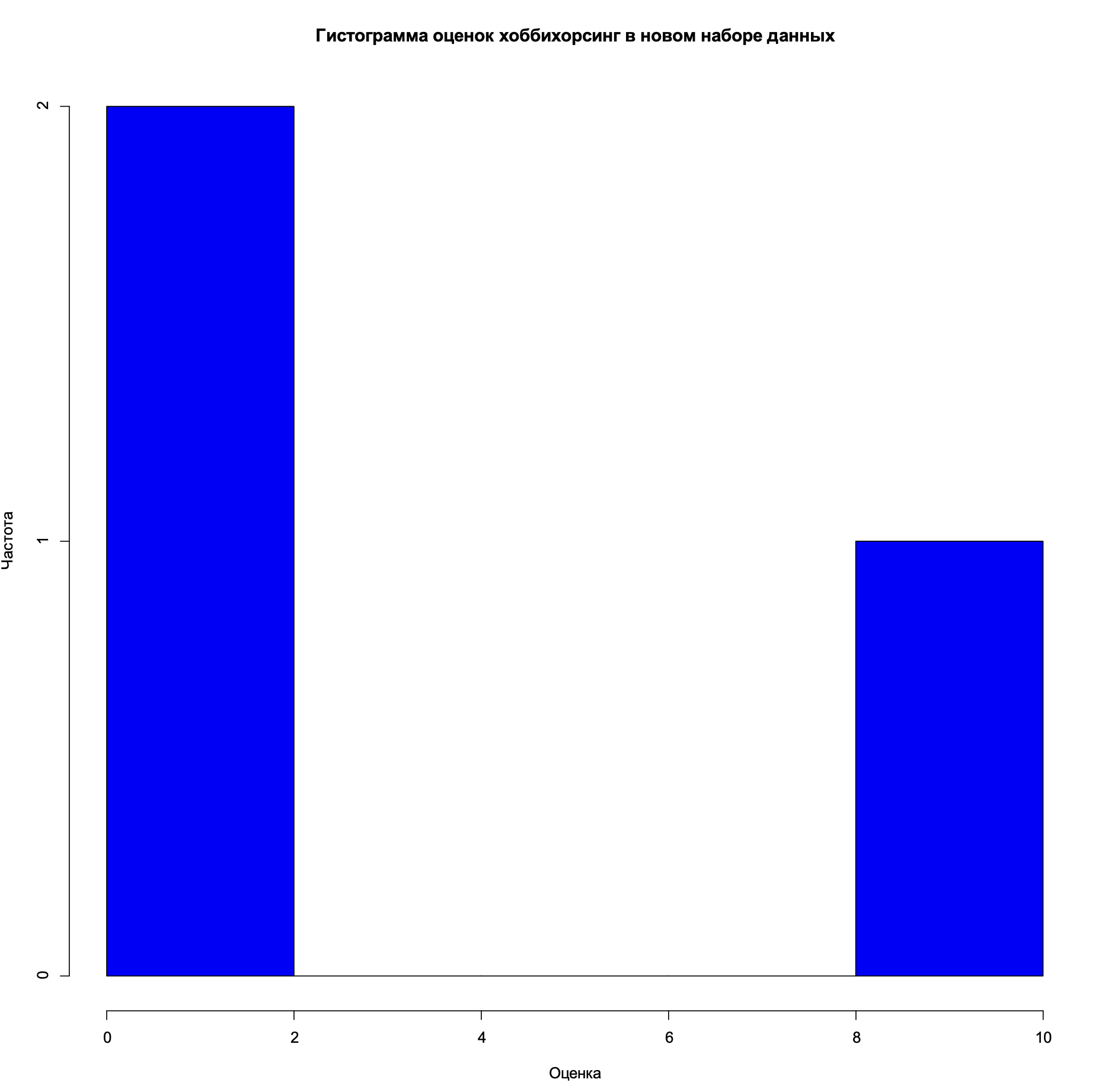


Рисунок 3 - Гистограмма оценок Red Bull

На рисунке 4 изображен боксплот на новых данных. хоббихорсинг теперь имеет более низкий минимум – количество высоких оценок снизилось. Изменения затронули и остальные виды спорта – они потеряли часть значений, что привело к их измененному распределению. Общий размах многих оценок уменьшился из-за фильтрации данных. Медианы изменились, так как остались только определённые респонденты.

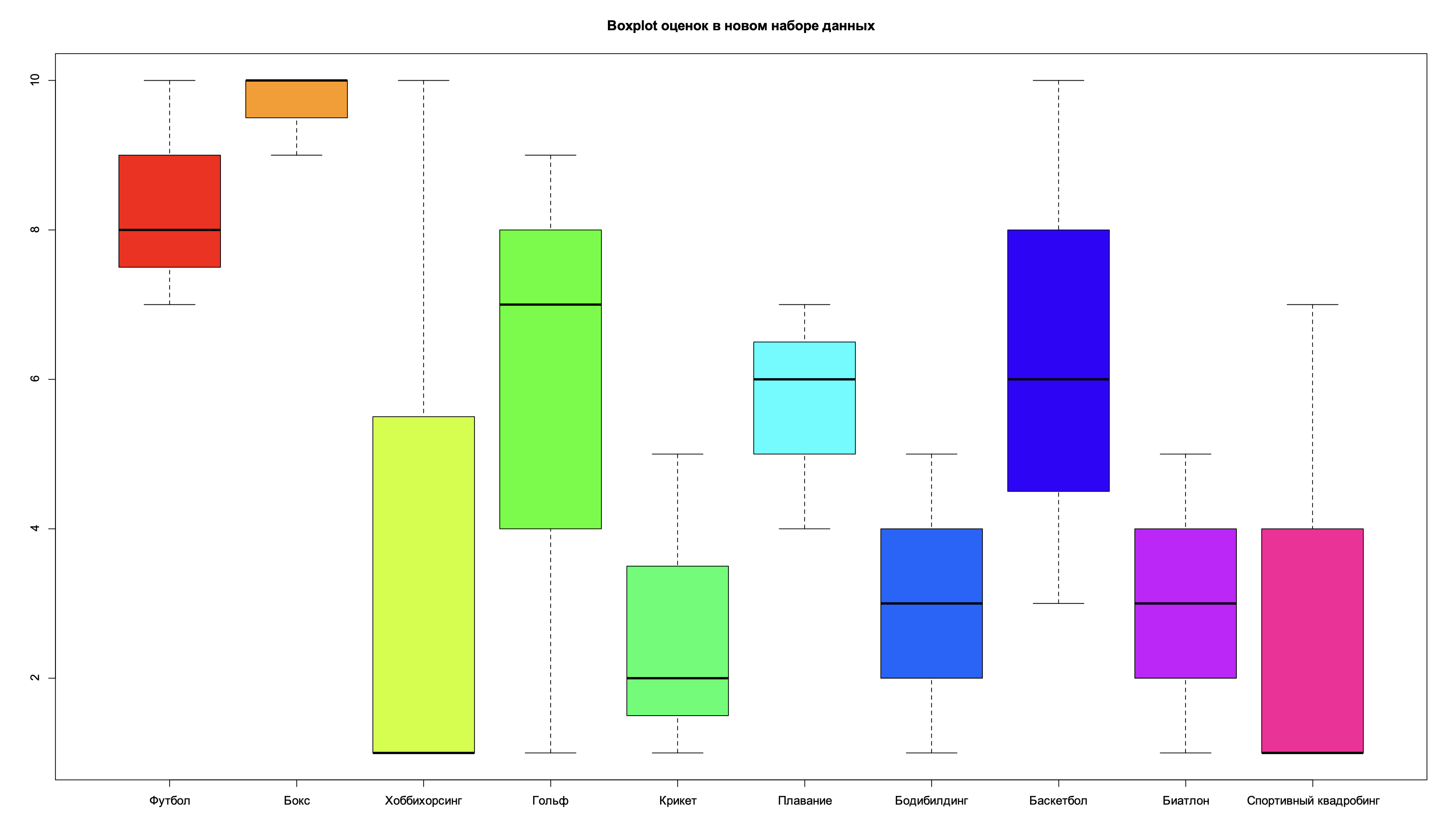


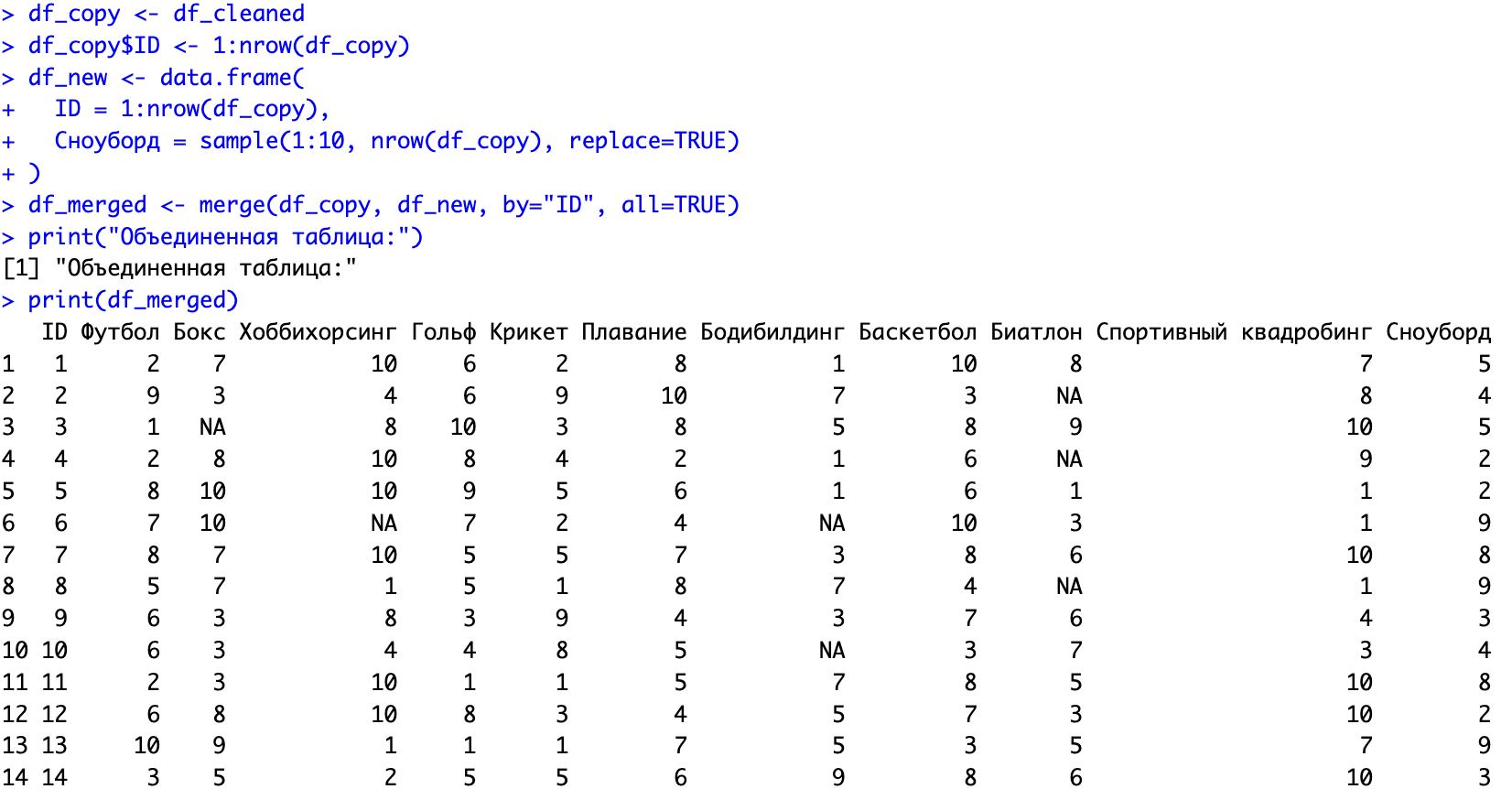
Рисунок 4 - Boxplot оценок в новом наборе данных

5. Продемонстрировать: слияние таблиц, добавление строк, исключение переменных,

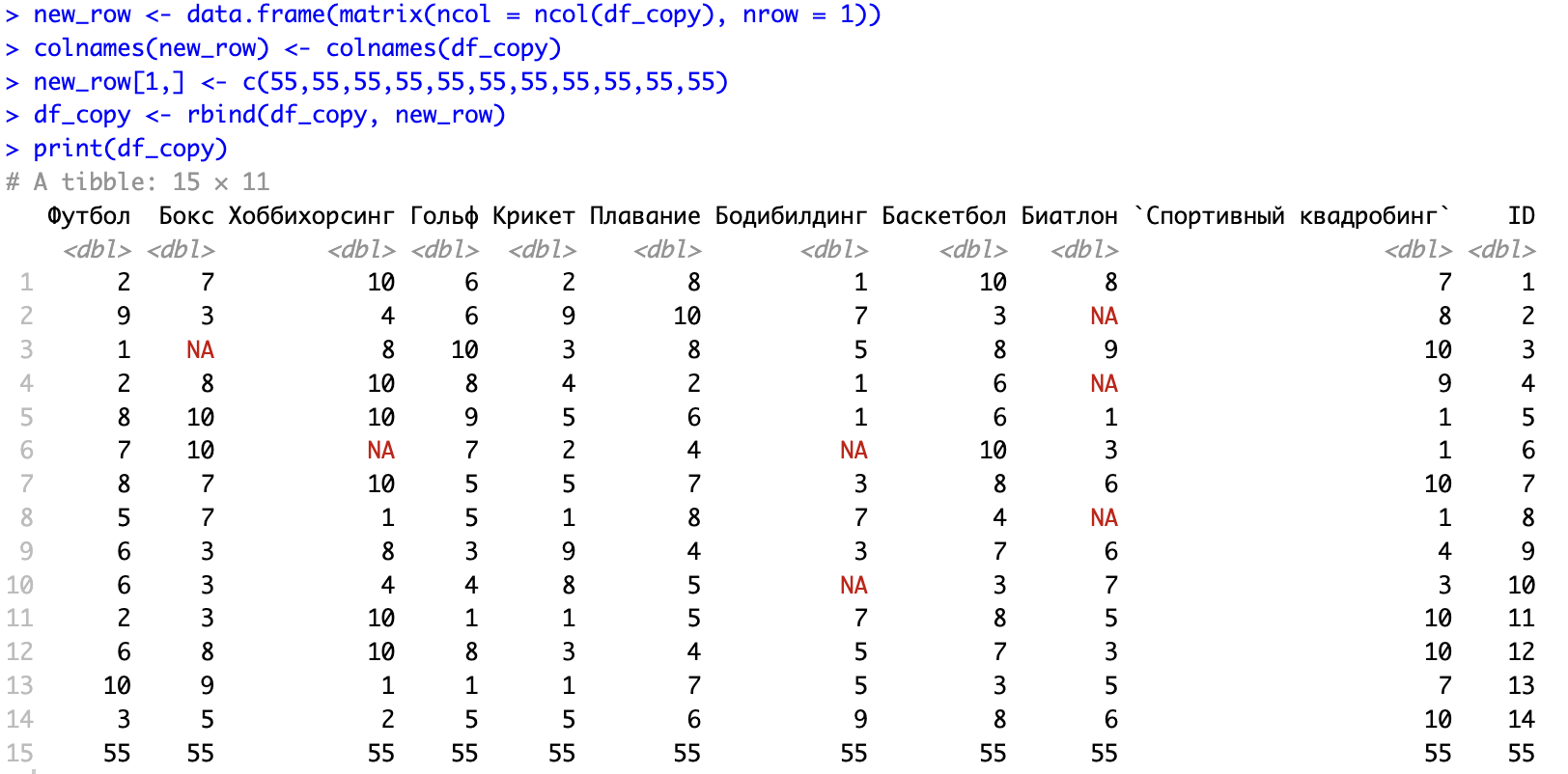
формирование части из целого набора данных - подмножество (subset), умение загрузить

данные их внешнего файла.

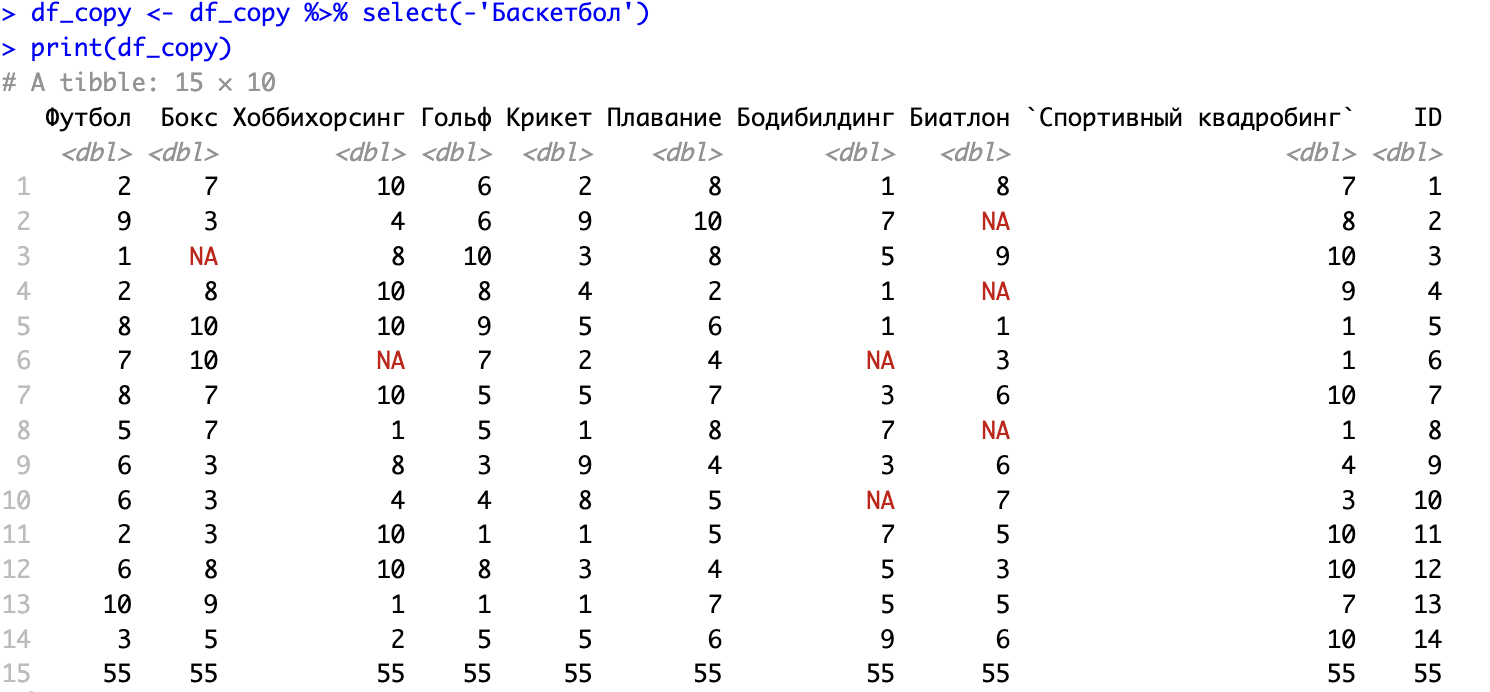
Слияние таблиц:



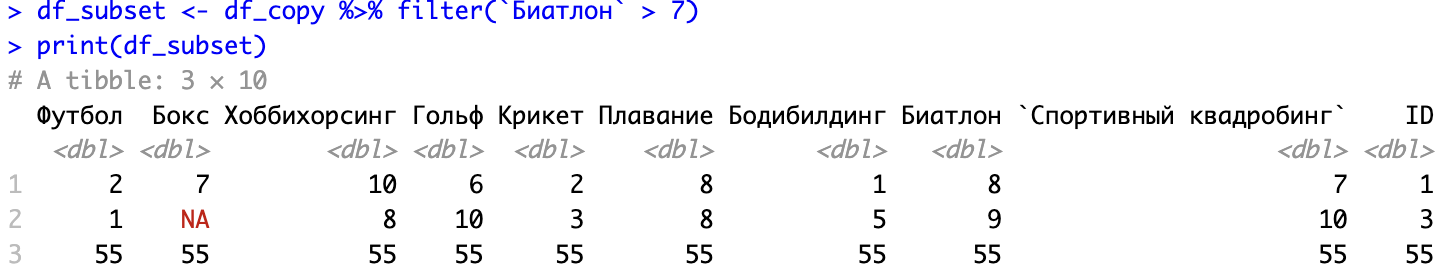
Добавление строк:



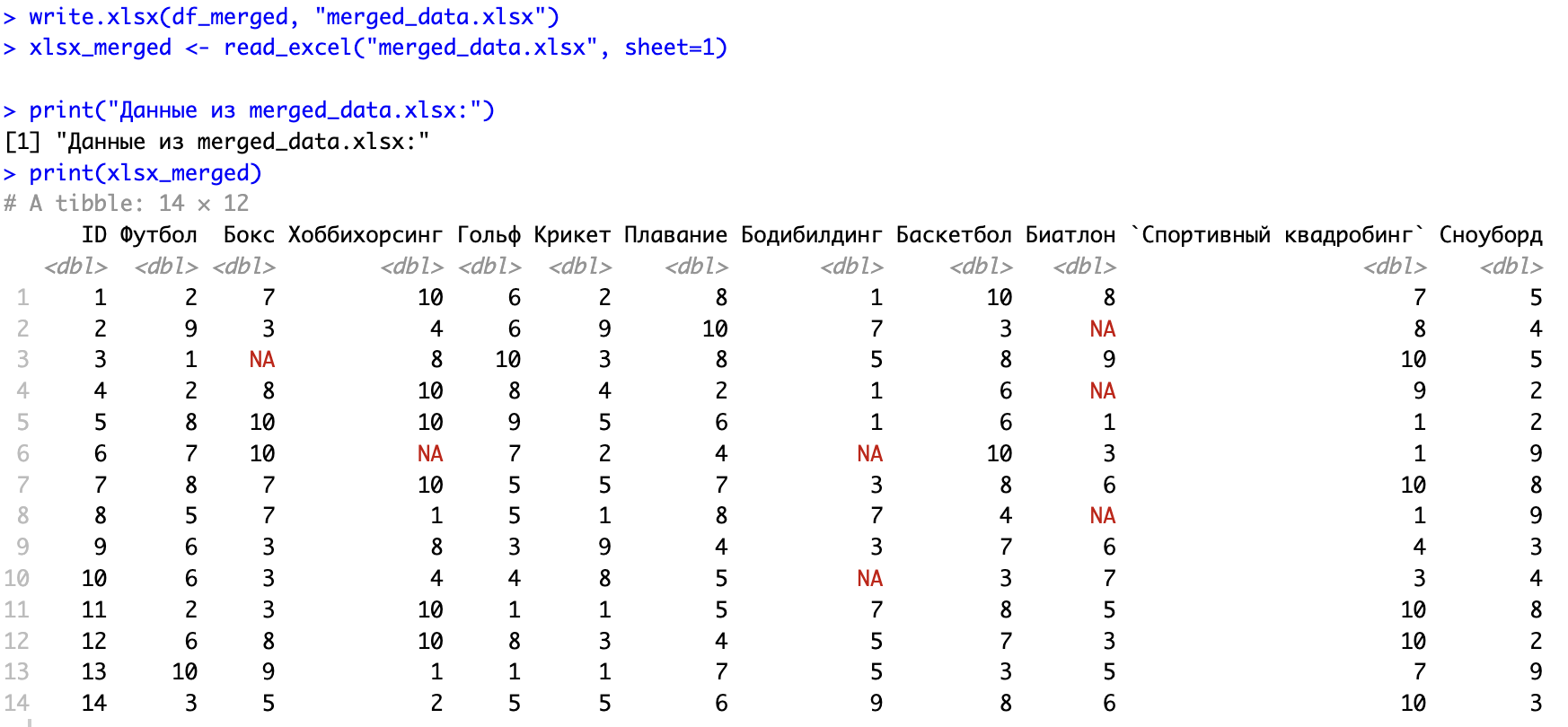
Исключение переменных:



Формирование подмножества:



Запись и чтение файла:



6. Результаты пояснить, сделать выводы (каждый график должен иметь

подрисуночную подпись – Рис.1. Название графика. И пояснения к графику).

**Вывод:** Научился получать элементарные характеристики наборов данных. Усвоил способы импорта данных. Научился трактовать некоторые элементарные способы графического представления наборов данных. Выполнил дескриптивный анализ данных.