|  |  |
| --- | --- |
| Gerb-BMSTU_01 | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

**Факультет «Информатика и системы управления»**

**Кафедра «Системы обработки информации и управления»**

Отчет по рубежному контролю №1

**«Технологии разведочного анализа и обработки данных»**

по дисциплине «Технологии машинного обучения»

Вариант №26

Выполнил:

студент группы ИУ5Ц-84Б   
Клеша Е.С.

подпись, дата

Проверил:

к.т.н., доц., Ю.Е. Гапанюк

подпись, дата

2024 г.

**СОДЕРЖАНИЕ ОТЧЕТА**

[1. Примечания: 4](#_Toc163693828)

[2. Дополнительные требования по группам: 4](#_Toc163693829)

[3. Листинг 5](#_Toc163693830)

[3.1. Импорт библиотек, загрузка данных 5](#_Toc163693831)

[3.2. Общее описание датасета 5](#_Toc163693832)

[3.3. Общее описание 5](#_Toc163693833)

[3.4. Корреляция признаков 6](#_Toc163693834)

[3.5. Визуальное исследование датасета 7](#_Toc163693835)

[3.6. Гистограмма для всех признаков 8](#_Toc163693836)

[3.7. Jointplot 10](#_Toc163693837)

[3.8. “Ящик с усами” 10](#_Toc163693838)

[3.8.1. По оси абсцисс 10](#_Toc163693839)

[3.8.2. По оси ординат 11](#_Toc163693840)

[3.9. Скрипичная диаграмма 11](#_Toc163693841)

# Примечания:

Если в Вашем наборе данных отсутствуют данные, необходимые для решения задачи, создайте их искусственно. Например, если отсутствуют категориальные признаки, создайте категориальный признак на основе числового. Если отсутствуют пропуски, замените на пропуски часть значений в одном или нескольких признаках.

Также Вы можете дополнительно использовать датасеты, содержащие необходимые данные, например использовать дополнительный датасет, содержащий пропуски.

# Дополнительные требования по группам:

1. Для студентов группы ИУ5-64Б, ИУ5Ц-84Б - для произвольной колонки данных построить график "Скрипичная диаграмма (violin plot)".

**Задача №4**

Для заданного набора данных постройте основные графики, входящие в этап разведочного анализа данных. В случае наличия пропусков в данных удалите строки или колонки, содержащие пропуски. Какие графики Вы построили и почему? Какие выводы о наборе данных Вы можете сделать на основании построенных графиков?

**Наборы данных:** https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load\_wine.html#sklearn.datasets.load\_wine

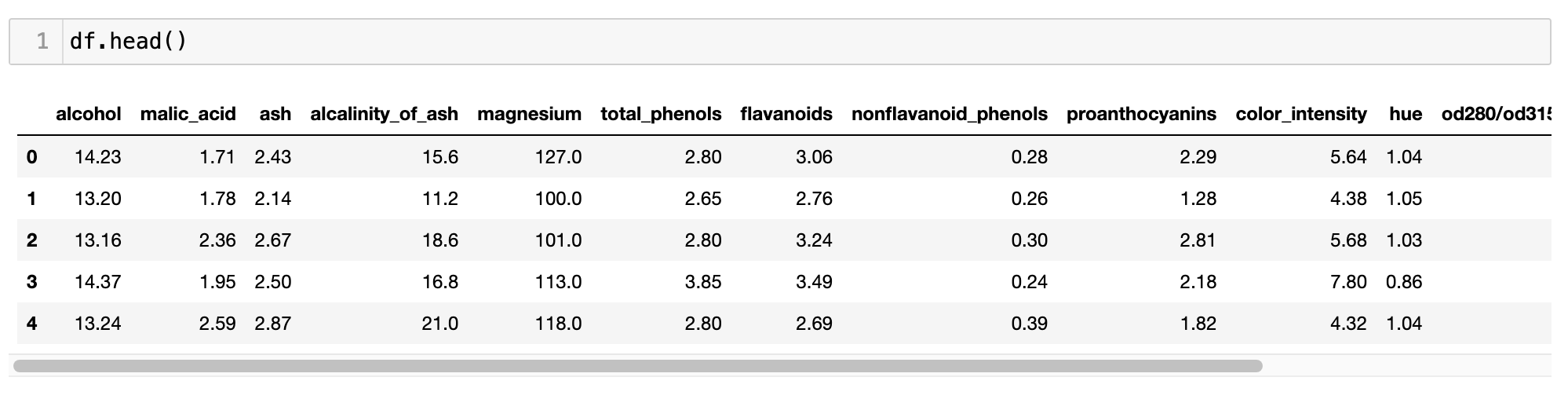
# Листинг

# Импорт библиотек, загрузка данных

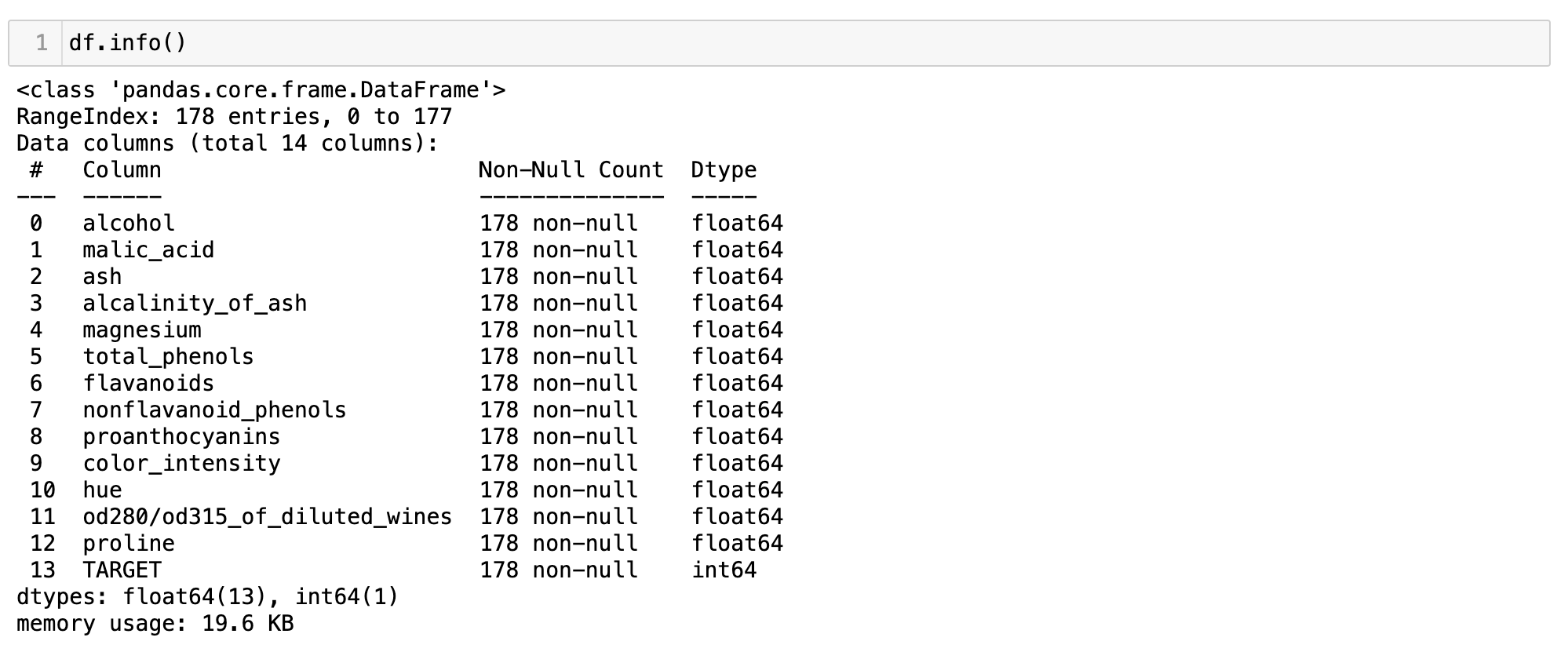


# Общее описание датасета

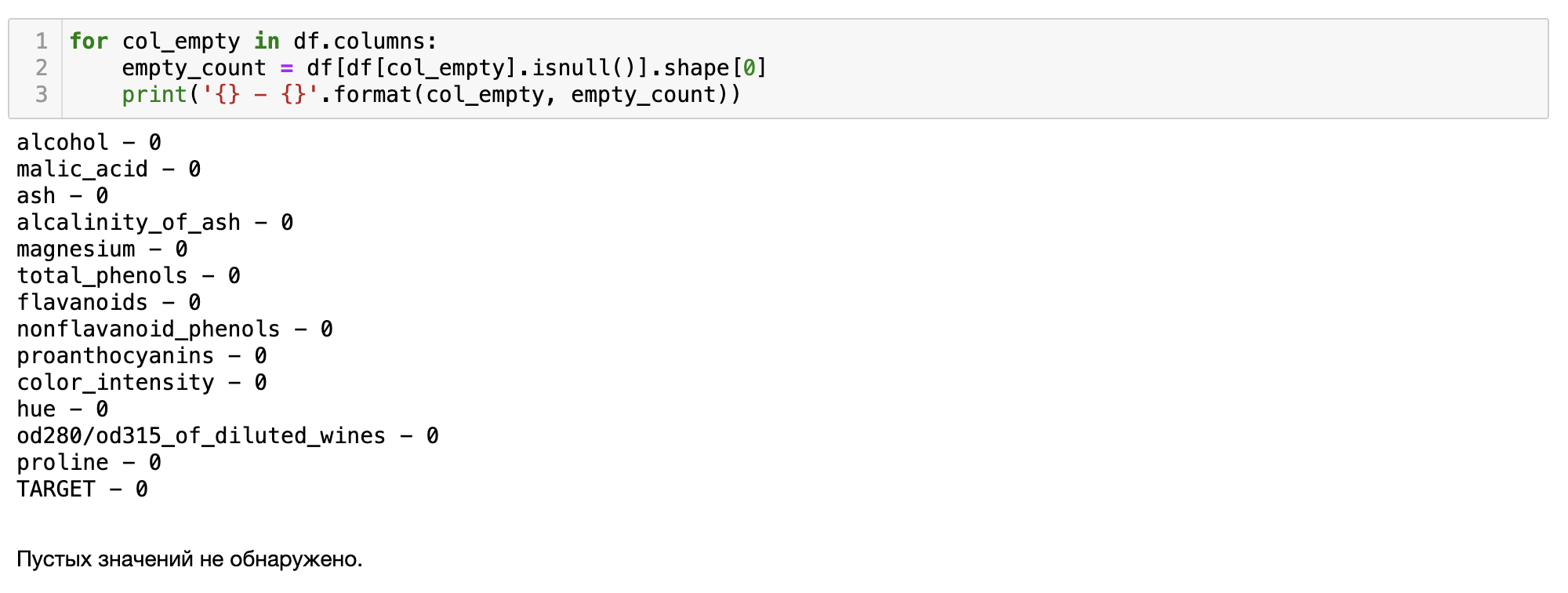
Первые пять строк датасета



# Общее описание

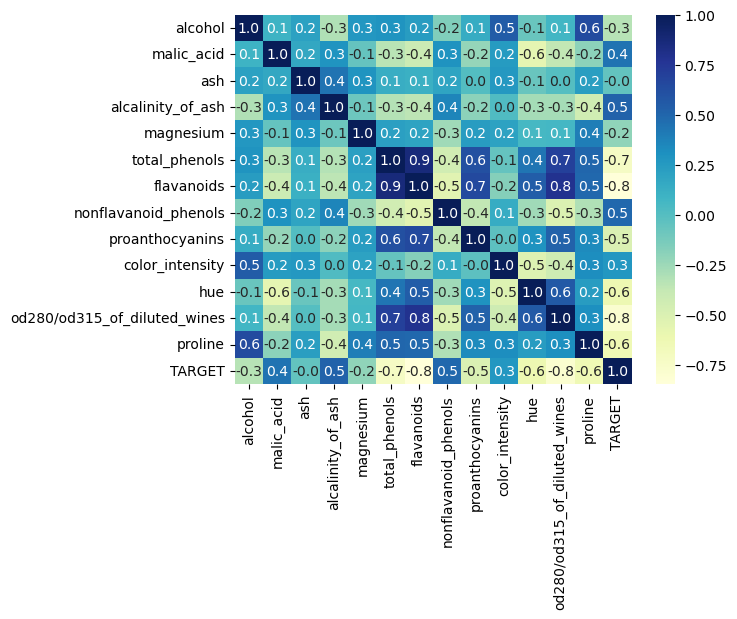


Проверим количество пустых значений

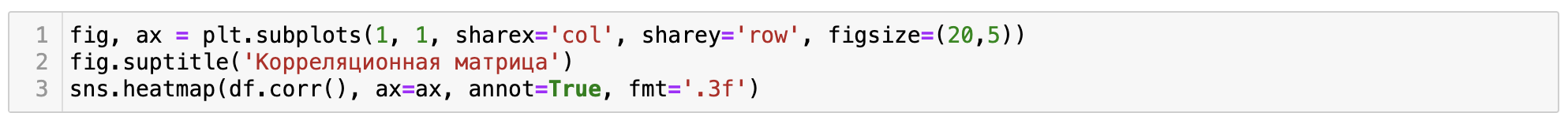


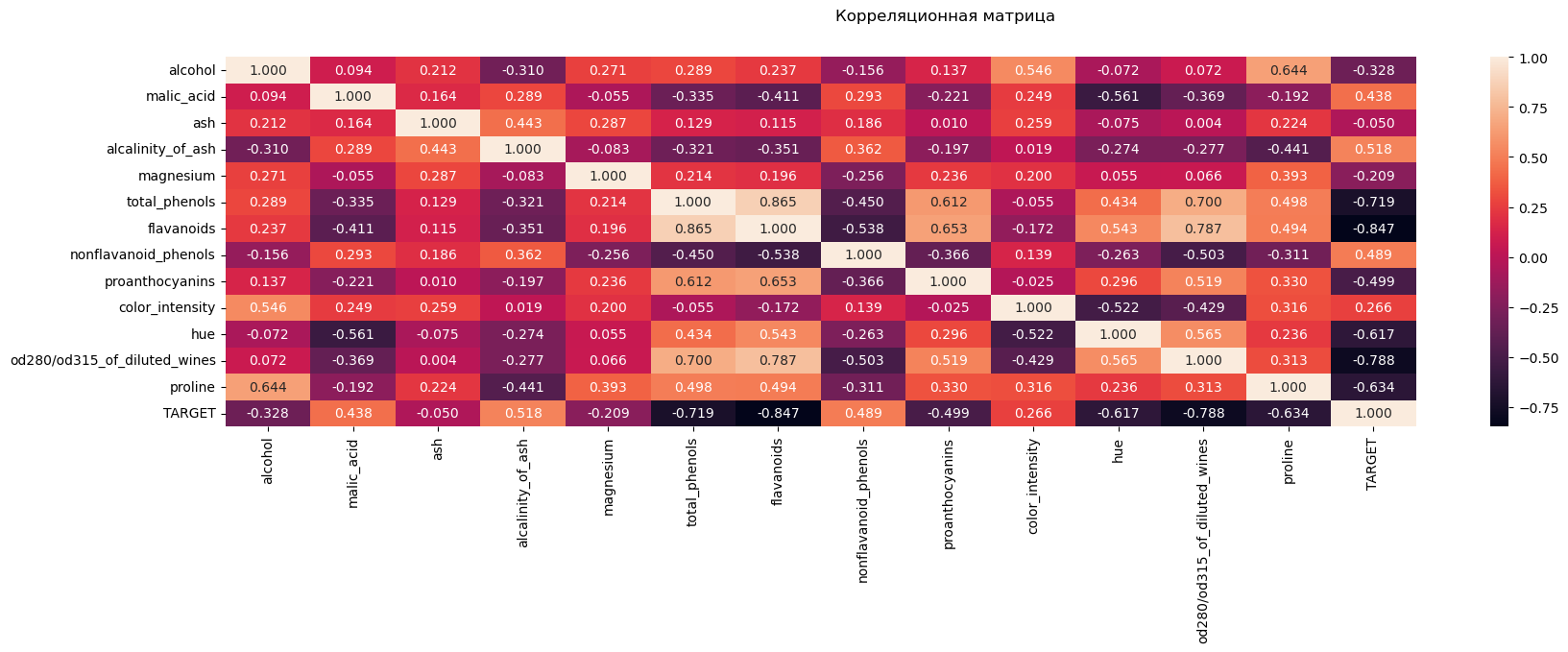
# Корреляция признаков





Наиболее сильную корреляцию имеют признаки total\_phenols и flavanoids. Это связано с тем, что флавониды относятся к классу полифенолов.



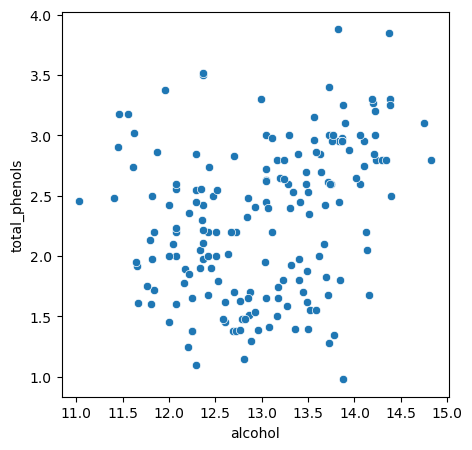


C целевым признаком TARGET сильнее всего коррелируют признаки "flavanoids", "od280/od315\_of\_diluted\_wines", "total\_phenols", "hue", "proline". Соответственно, их стоит учитывать для более информативного построения модели машинного обучения.

# Визуальное исследование датасета

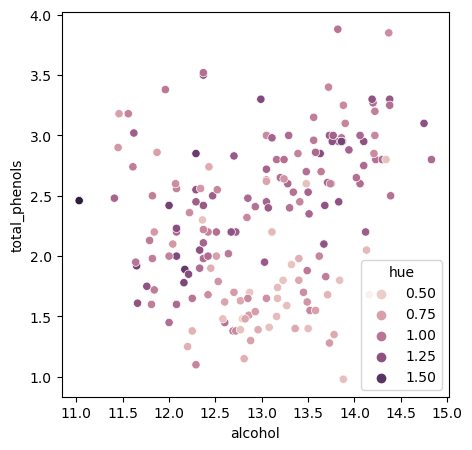
Диаграмма рассеивания для признаков total\_phenols и alcohol





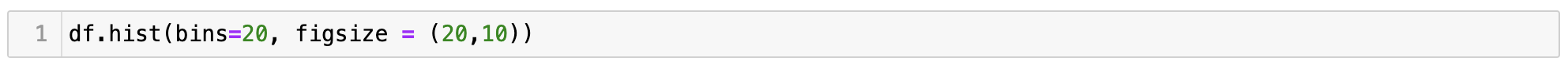
Данная диаграмма показывает количество фенолов в каждом проценте вина.

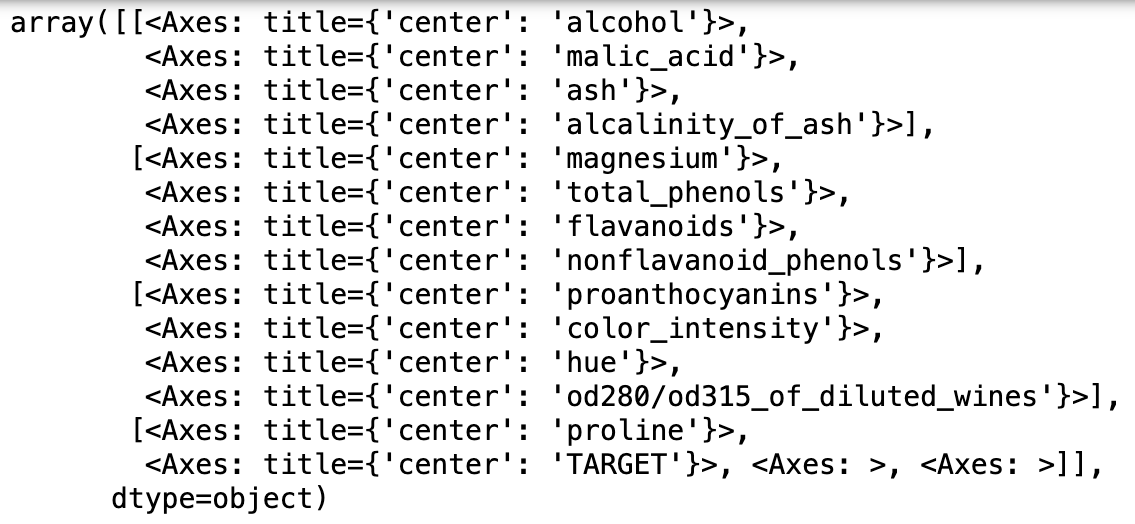


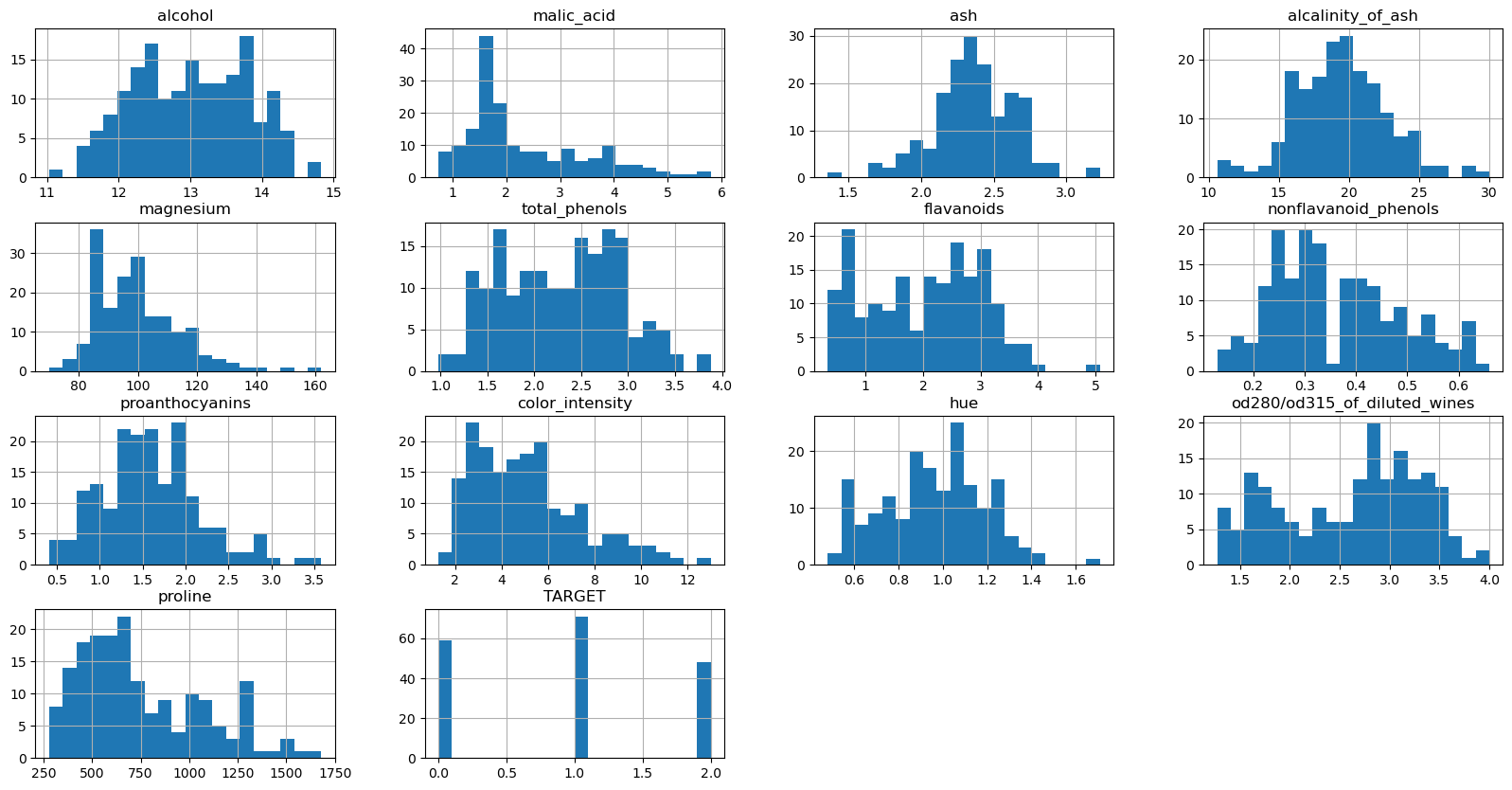


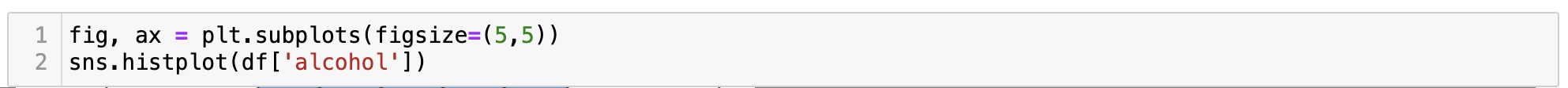
Такая же диаграмма показывает количество фенолов в каждом проценте вина, но еще добавили "hue", т.е. в каждой точке можем рассмотреть оттеннок конкретного вина.

# Гистограмма для всех признаков

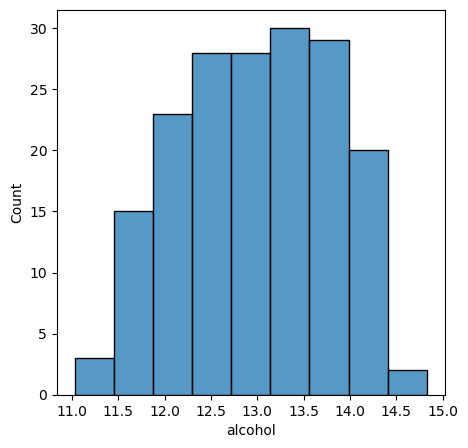






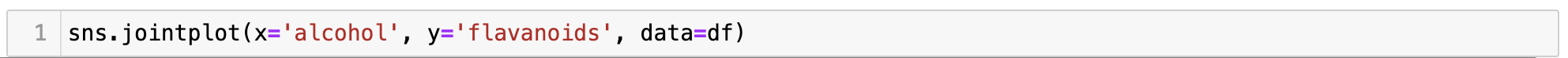




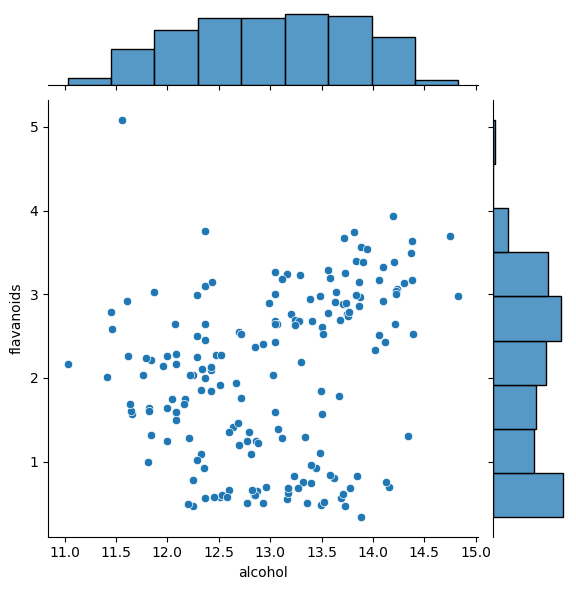


Данная гистограмма показывает наибольшее количество процента алкоголя в вине.

# Jointplot





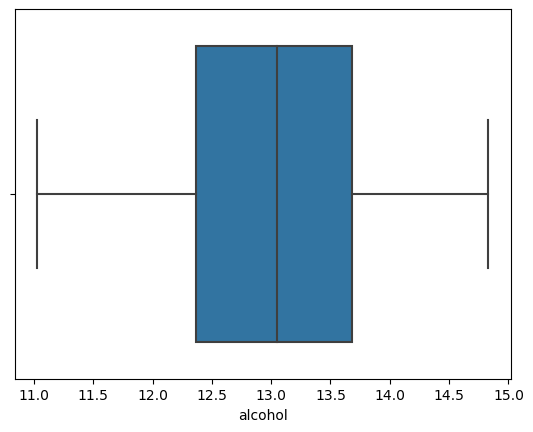


Комбинация гистограмм и диаграмм рассеивания.

# “Ящик с усами”

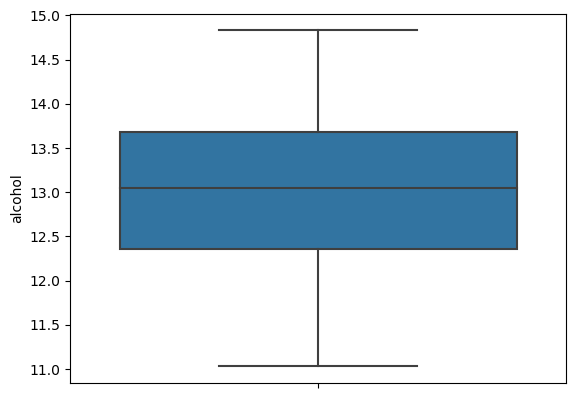
# По оси абсцисс





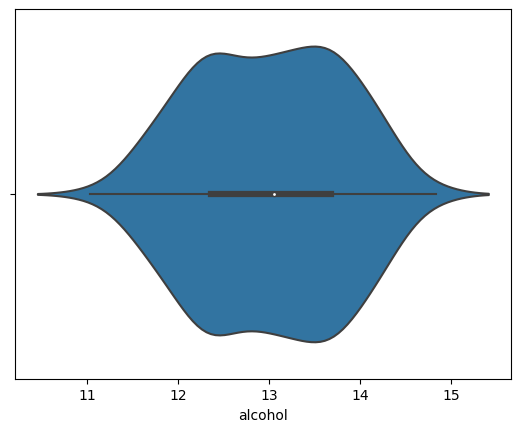
# По оси ординат





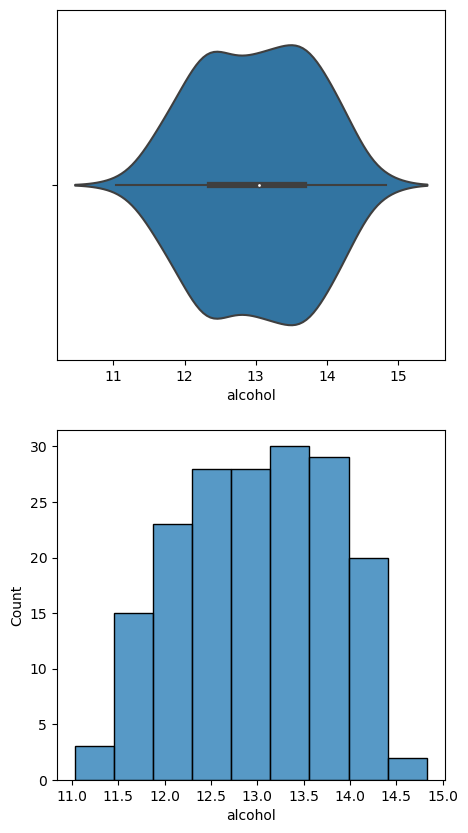
# Скрипичная диаграмма





Скрипичная диаграмма показывает распределение плотности по краям диаграммы.





Из приведенных графиков видно, что скрипичная диаграмма действительно показывает распределение плотности.