

**Московский государственный технический
университет им. Н. Э. Баумана**

Курс «Технологии машинного обучения»

Отчёт по лабораторной работе №1

Выполнил:
Здобняков Ф. А.
группа ИУ5-64Б

Проверил:
Гапанюк Ю.Е.

Дата: 07.04.25

Дата:

Подпись:

Подпись:

Москва, 2025 г.

Цель лабораторной работы: изучение различных методов визуализация данных.

Краткое описание. Построение основных графиков, входящих в этап разведочного анализа данных.

Рекомендуемые инструментальные средства можно посмотреть [здесь](#).

Задание:

- Выбрать набор данных (датасет). Вы можете найти список свободно распространяемых датасетов [здесь](#).
- Для первой лабораторной работы рекомендуется использовать датасет без пропусков в данных, например из [Scikit-learn](#).
- Пример преобразования датасетов Scikit-learn в Pandas Dataframe можно посмотреть [здесь](#).

Для лабораторных работ не рекомендуется выбирать датасеты большого размера.

- Создать ноутбук, который содержит следующие разделы:
 1. Текстовое описание выбранного Вами набора данных.
 2. Основные характеристики датасета.
 3. Визуальное исследование датасета.
 4. Информация о корреляции признаков.
- Сформировать отчет и разместить его в своем репозитории на github.

Ход выполнения:

import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.set(style="ticks")

data = pd.read_csv('wine.data.txt', sep=",")

data.head()

	Класс	Алкоголь	Яблочная кислота	Пепел	Щелочность золы	Магний	Всего фенолов	Флаваноиды	Нефлаваноидные фенолы	Проантоцианы	Интенсивность цвета	Оттенок	OD280/OD315 разбавленных вин	Пролин
0	1	14.23	1.71	2.43	15.6	127	2.80	3.06	0.28	2.29	5.64	1.04	3.92	1065
1	1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28	4.38	1.05	3.40	1050
2	1	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30	2.81	5.68	1.03	3.17	1185
3	1	14.37	1.95	2.50	16.8	113	3.85	3.49	0.24	2.18	7.80	0.86	3.45	1480
4	1	13.24	2.59	2.87	21.0	118	2.80	2.69	0.39	1.82	4.32	1.04	2.93	735

data.shape

(178, 14)

data.columns

Index(['Класс', 'Алкоголь', 'Яблочная кислота', 'Пепел', 'Щелочность золы', 'Магний', 'Всего фенолов', 'Флаваноиды', 'Нефлаваноидные фенолы', 'Проантоцианы', 'Интенсивность цвета', 'Оттенок', 'OD280/OD315 разбавленных вин', 'Пролин'], dtype='object')

data.dtypes

Класс	int64
Алкоголь	float64
Яблочная кислота	float64
Пепел	float64
Щелочность золы	float64
Магний	int64
Всего фенолов	float64
Флаваноиды	float64
Нефлаваноидные фенолы	float64
Проантоцианы	float64
Интенсивность цвета	float64
Оттенок	float64
OD280/OD315 разбавленных вин	float64
Пролин	int64
dtype:	object

for col in data.columns:
 # Количество пустых значений – все значения заполнены
 temp_null_count = data[data[col].isnull()].shape[0]
 print('{} - {}'.format(col, temp_null_count))

Класс - 0
Алкоголь - 0
Яблочная кислота - 0
Пепел - 0
Щелочность золы - 0
Магний - 0
Всего фенолов - 0

Spaces: 4 Ячейка 1 из 27 Go Live

Лаб1Вино).ipynb

+ Код

+ Markdown

▶ Выполнить все

≡ Очистить все выходные данные

≡ Структура

...

Выбрать ядро

Флаванои́ды - 0

Нефлаваноидные фенолы - 0

Проантоцианы - 0

Интенсивность цвета - 0

Оттенок - 0

00280/00315 разбавленных вин - 0

Пролин - 0

▶

▼

[12]

data.describe()

Python

...

	Класс	Алкоголь	Яблочная кислота	Пепел	Щелочность золы	Магний	Всего фенолов	Флаванои́ды	Нефлаваноидные фенолы	Проантоцианы	Интенсивность цвета	Оттен
count	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.0000
mean	1.938202	13.000618	2.336348	2.366517	19.494944	99.741573	2.295112	2.029270	0.361854	1.590899	5.058090	0.9574
std	0.775035	0.811827	1.117146	0.274344	3.339564	14.282484	0.625851	0.998859	0.124453	0.572359	2.318286	0.228E
min	1.000000	11.030000	0.740000	1.360000	10.600000	70.000000	0.980000	0.340000	0.130000	0.410000	1.280000	0.4800
25%	1.000000	12.362500	1.602500	2.210000	17.200000	88.000000	1.742500	1.205000	0.270000	1.250000	3.220000	0.7825
50%	2.000000	13.050000	1.865000	2.360000	19.500000	98.000000	2.355000	2.135000	0.340000	1.555000	4.690000	0.9650
75%	3.000000	13.677500	3.082500	2.557500	21.500000	107.000000	2.800000	2.875000	0.437500	1.950000	6.200000	1.1200
max	3.000000	14.830000	5.800000	3.230000	30.000000	162.000000	3.880000	5.080000	0.660000	3.580000	13.000000	1.7100

[15]

data['Класс'].unique()

Python

... array([1, 2, 3])

[21]

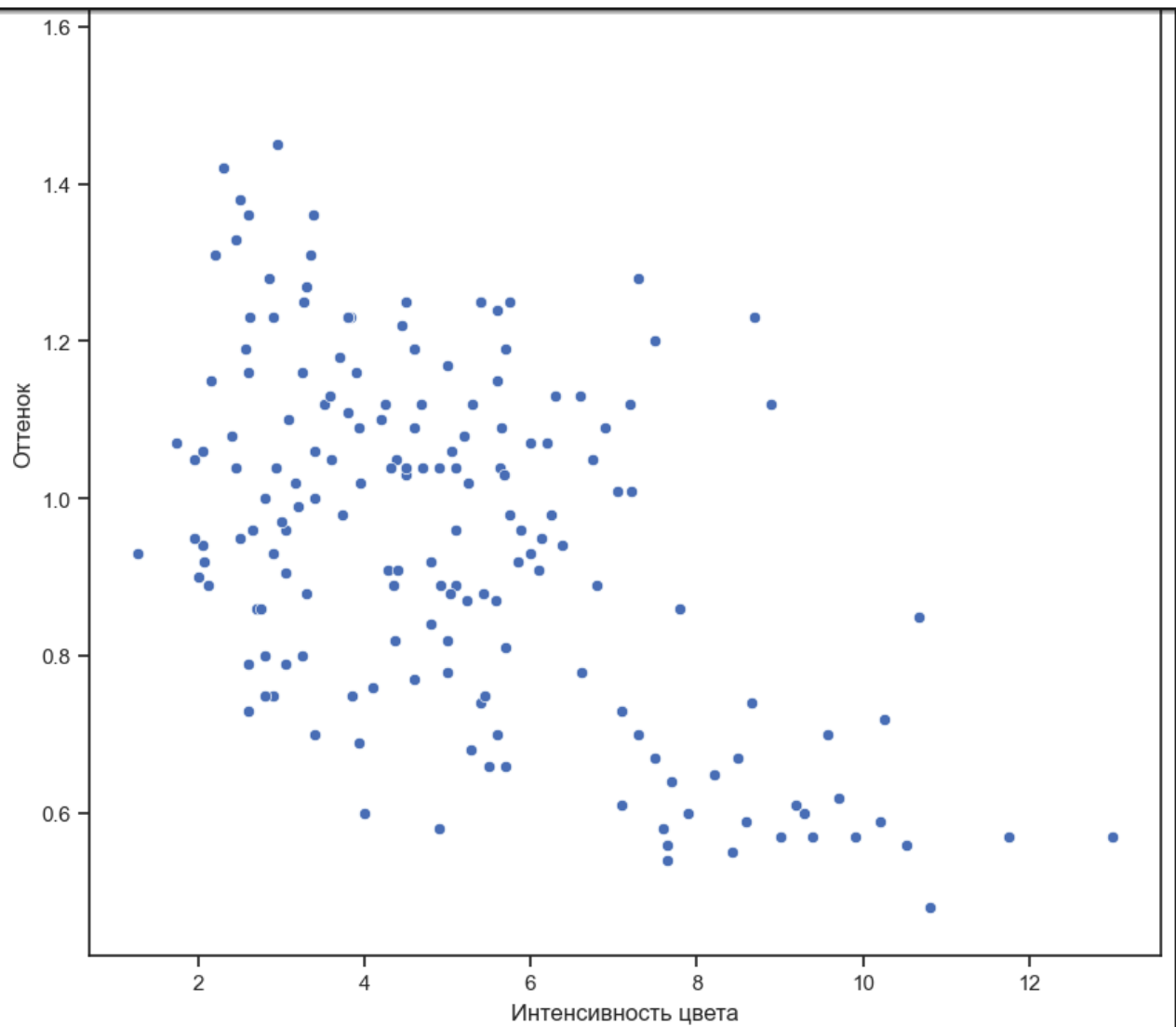
fig, ax = plt.subplots(figsize=(10,10))
sns.scatterplot(ax=ax, x='Интенсивность цвета', y='Оттенок', data=data)

Python

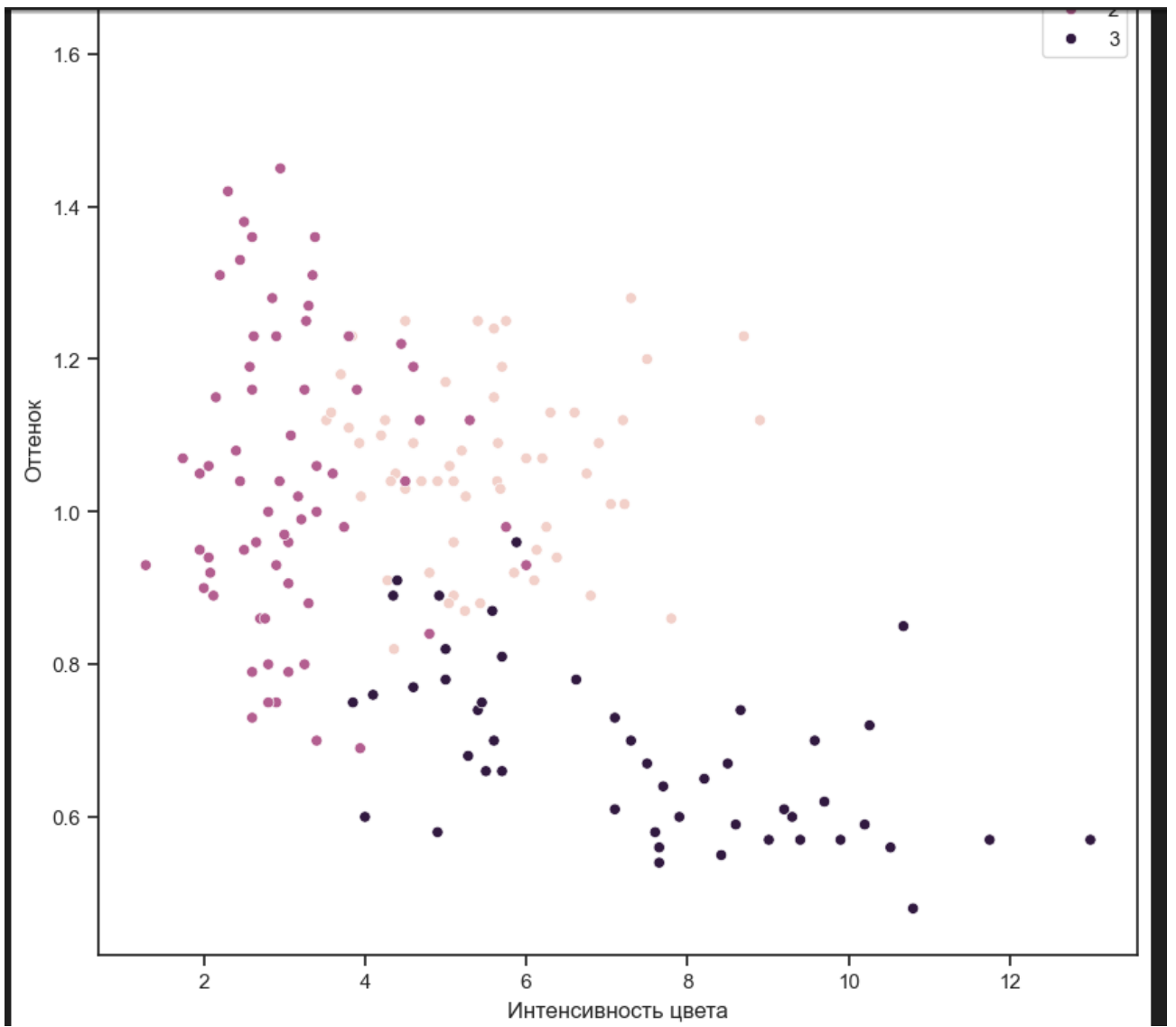
... <Axes: xlabel='Интенсивность цвета', ylabel='Оттенок'>

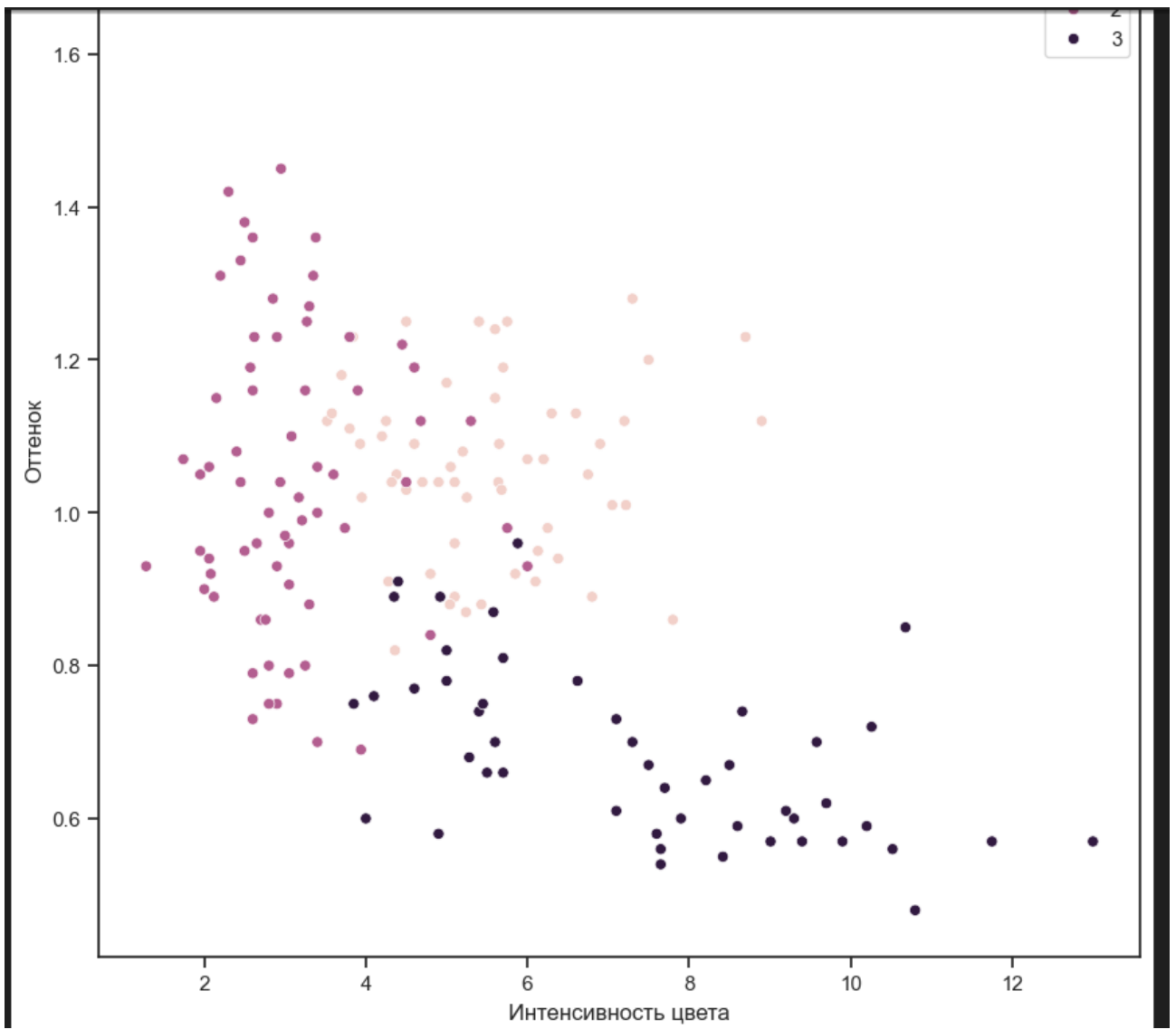
...

Spaces: 4 Ячейка 1 из 27 Go Live

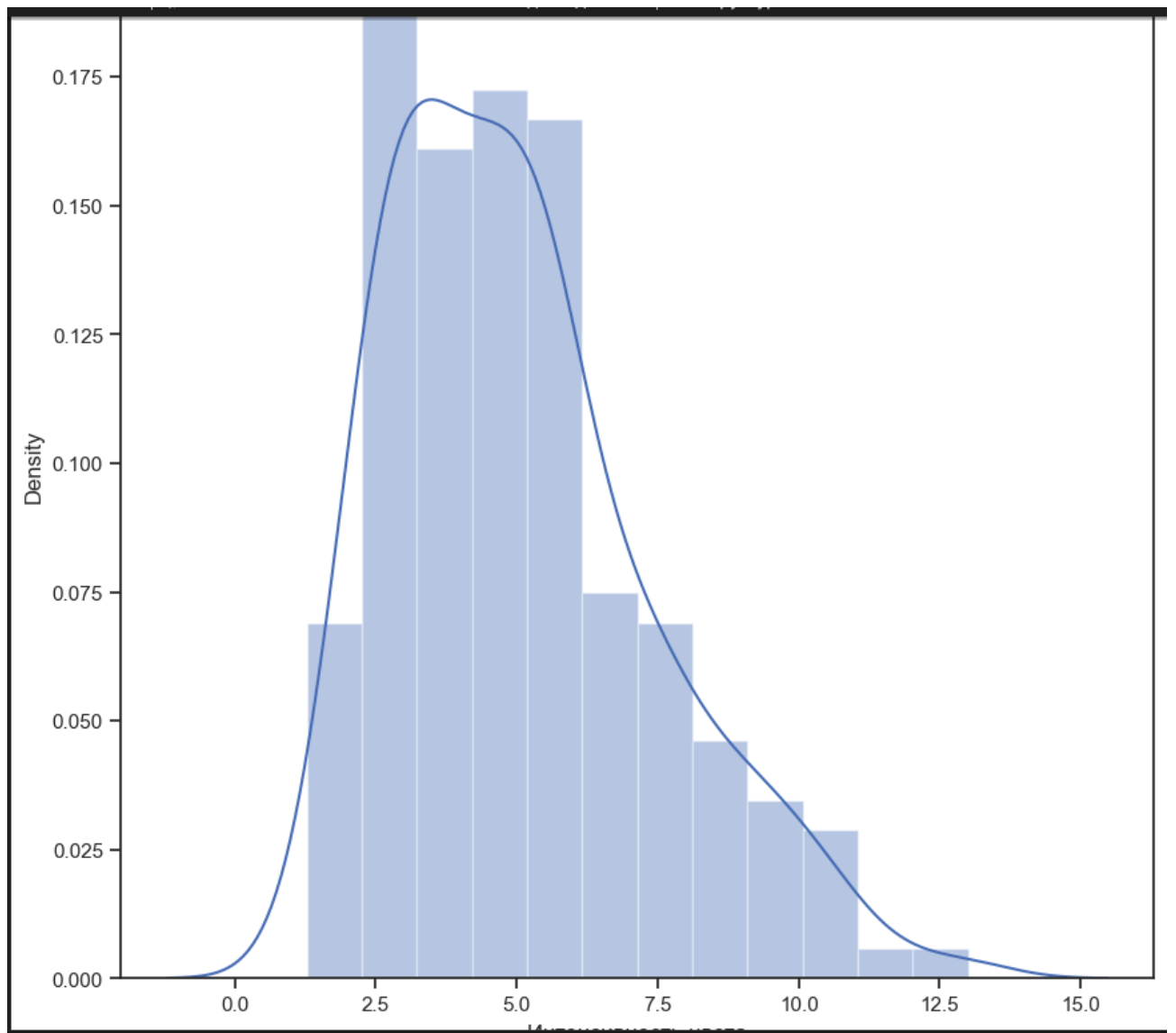


```
fig, ax = plt.subplots(figsize=(10,10))  
sns.scatterplot(ax=ax, x='Интенсивность цвета', y='Оттенок', data=data, hue='Класс')
```



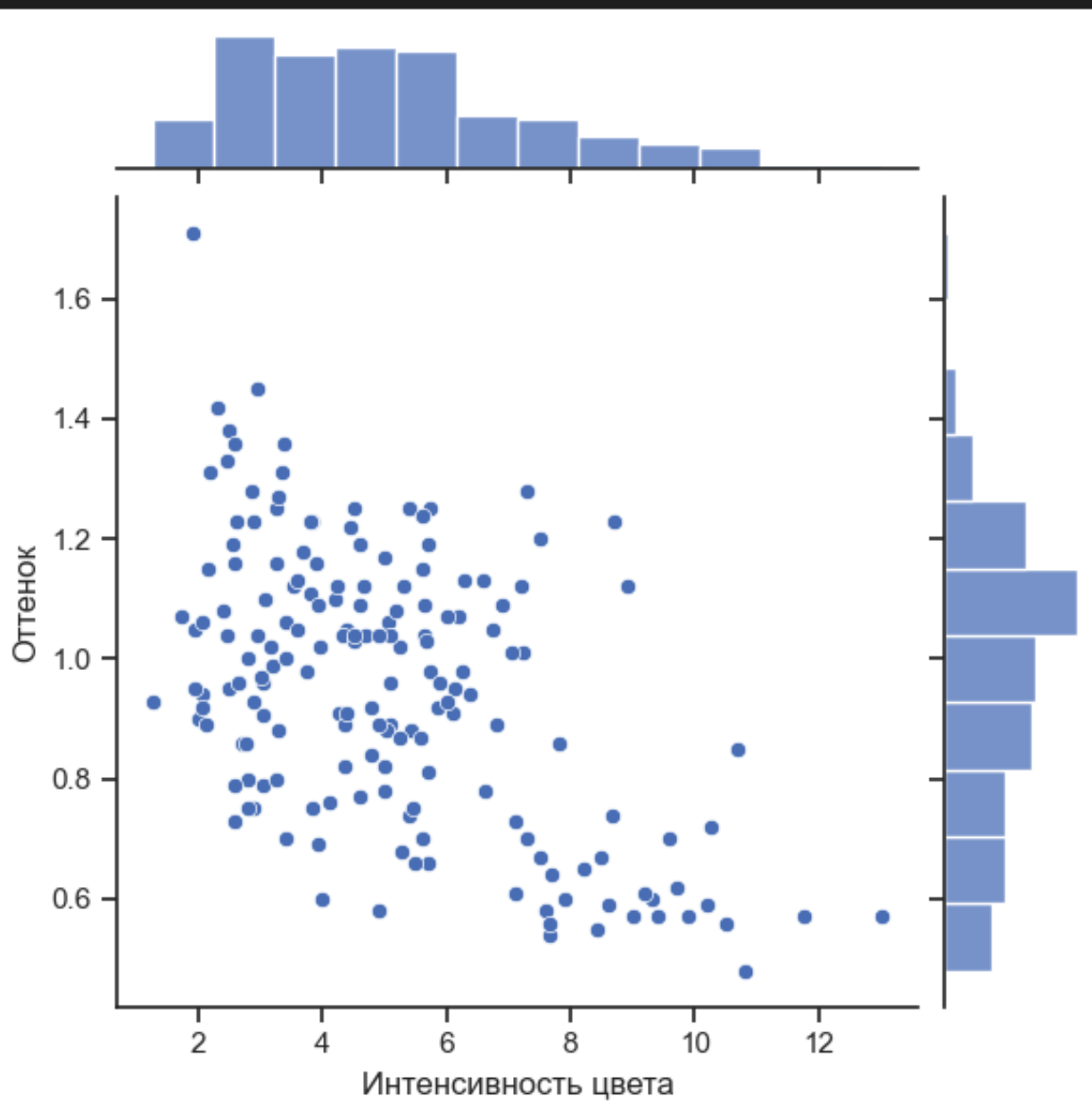


```
fig, ax = plt.subplots(figsize=(10,10))  
sns.distplot(data['Интенсивность цвета'])
```



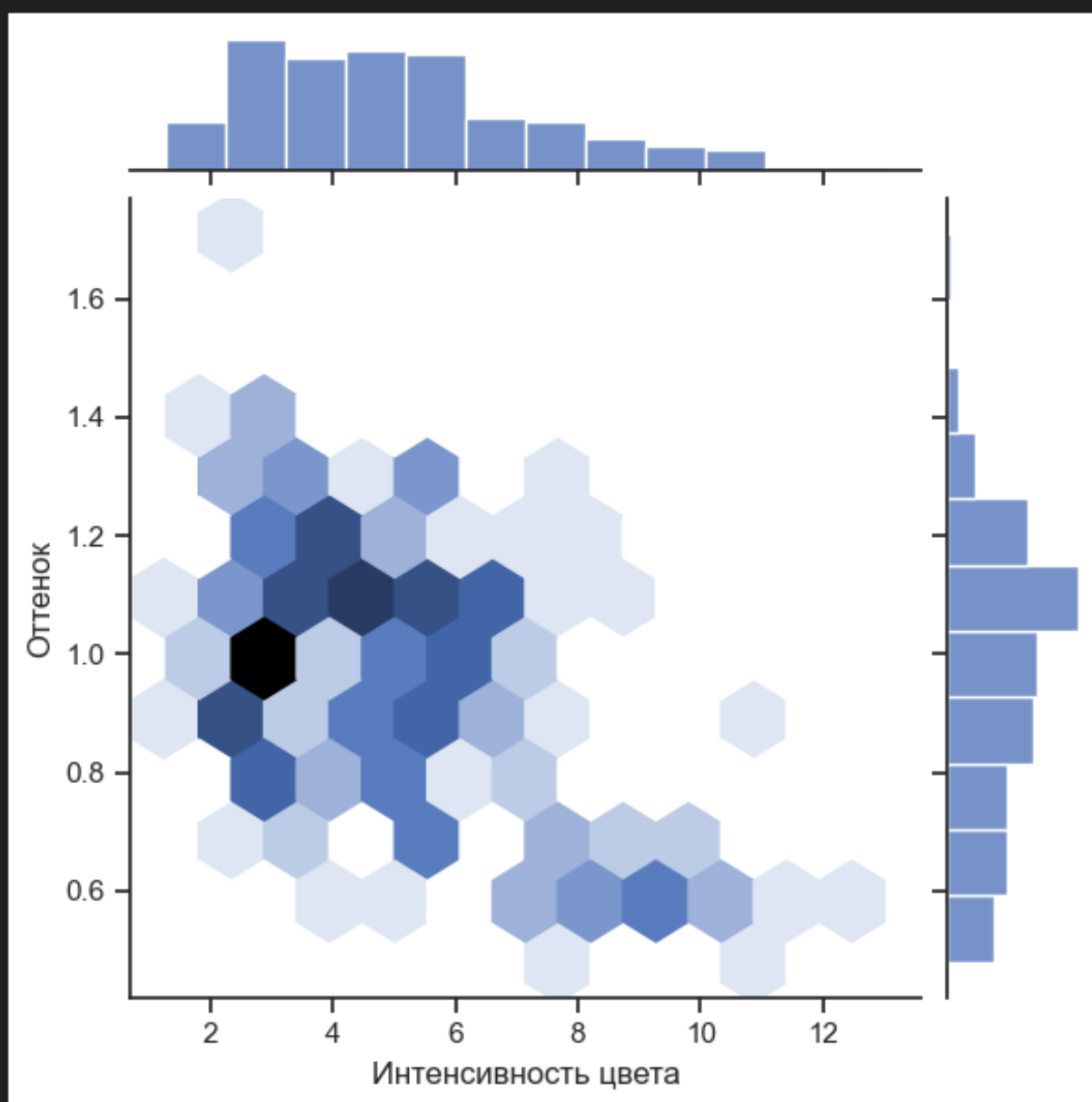

```
sns.jointplot(x='Интенсивность цвета', y='Оттенок', data=data)
```

<seaborn.axisgrid.JointGrid at 0x1163eef00>



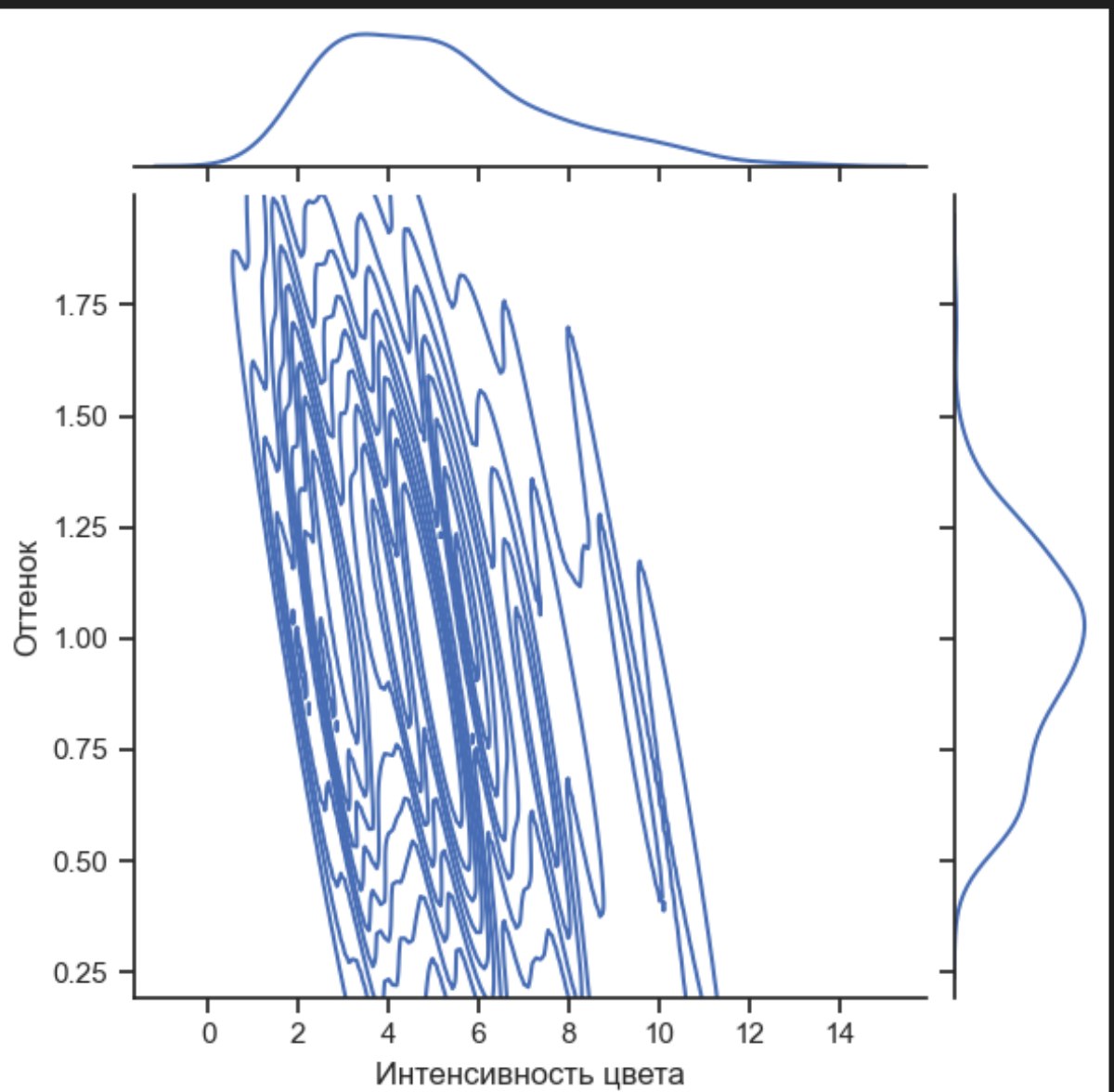
```
sns.jointplot(x='Интенсивность цвета', y='Оттенок', data=data, kind="hex")
```

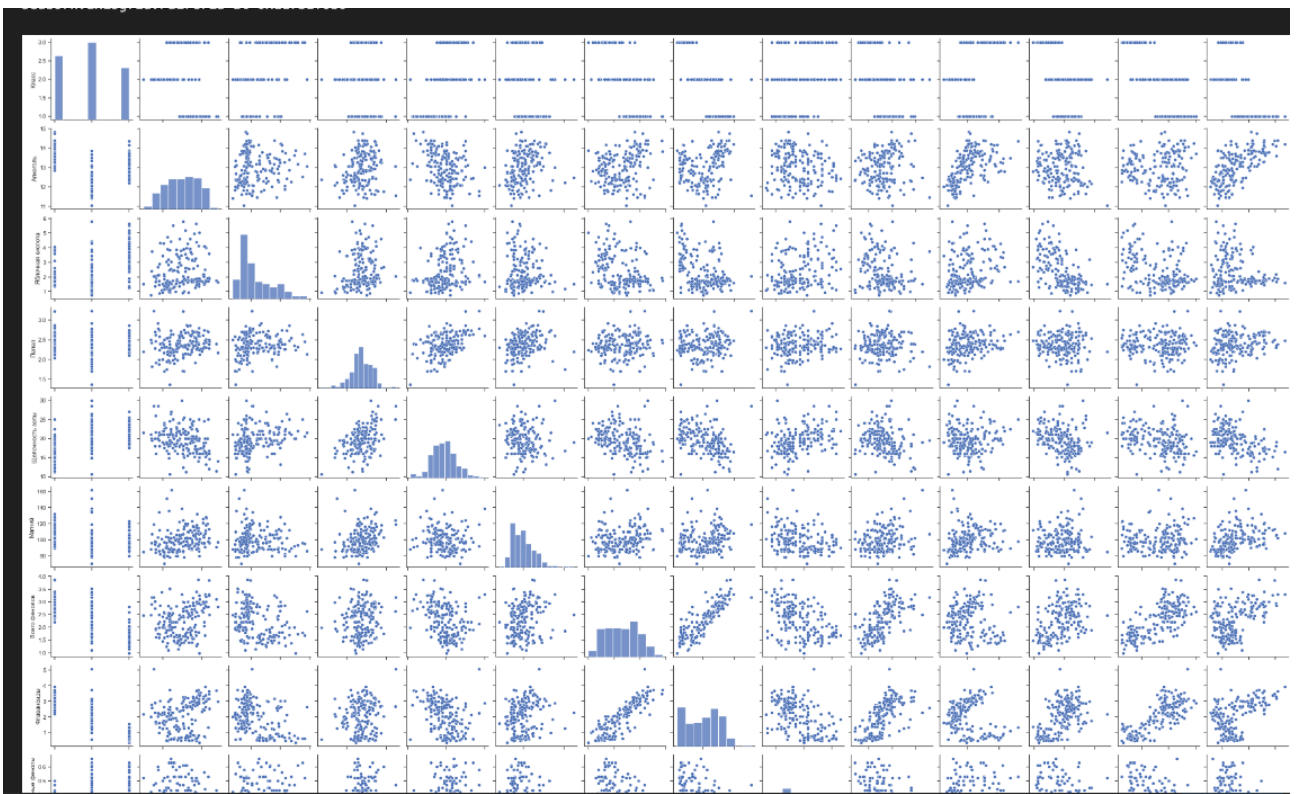
<seaborn.axisgrid.JointGrid at 0x11798e420>



```
sns.jointplot(x='Интенсивность цвета', y='Оттенок', data=data, kind="kde")
```

<seaborn.axisgrid.JointGrid at 0x117b1c3e0>

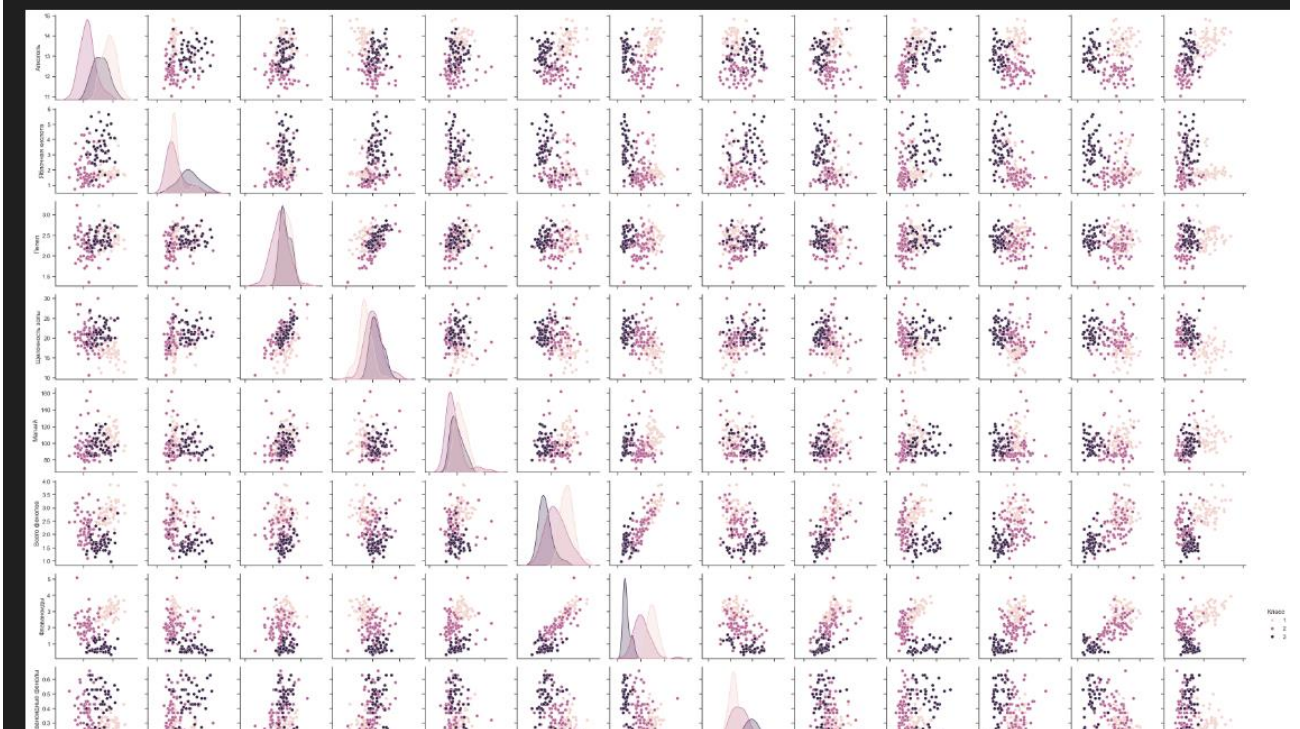




```
sns.pairplot(data, hue="Knacc")
```

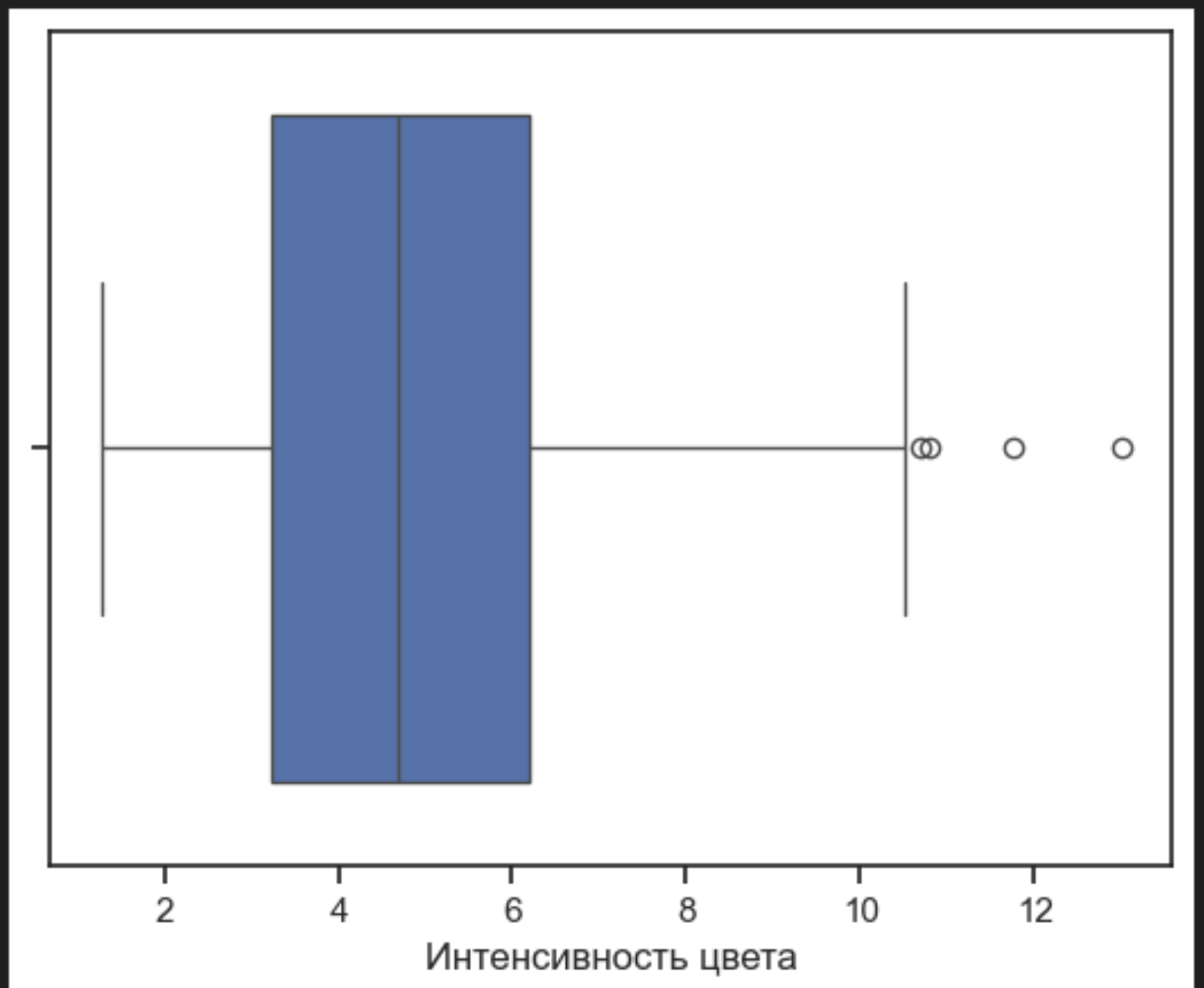
Pyth

<seaborn.axisgrid.PairGrid at 0x11a4663f0>



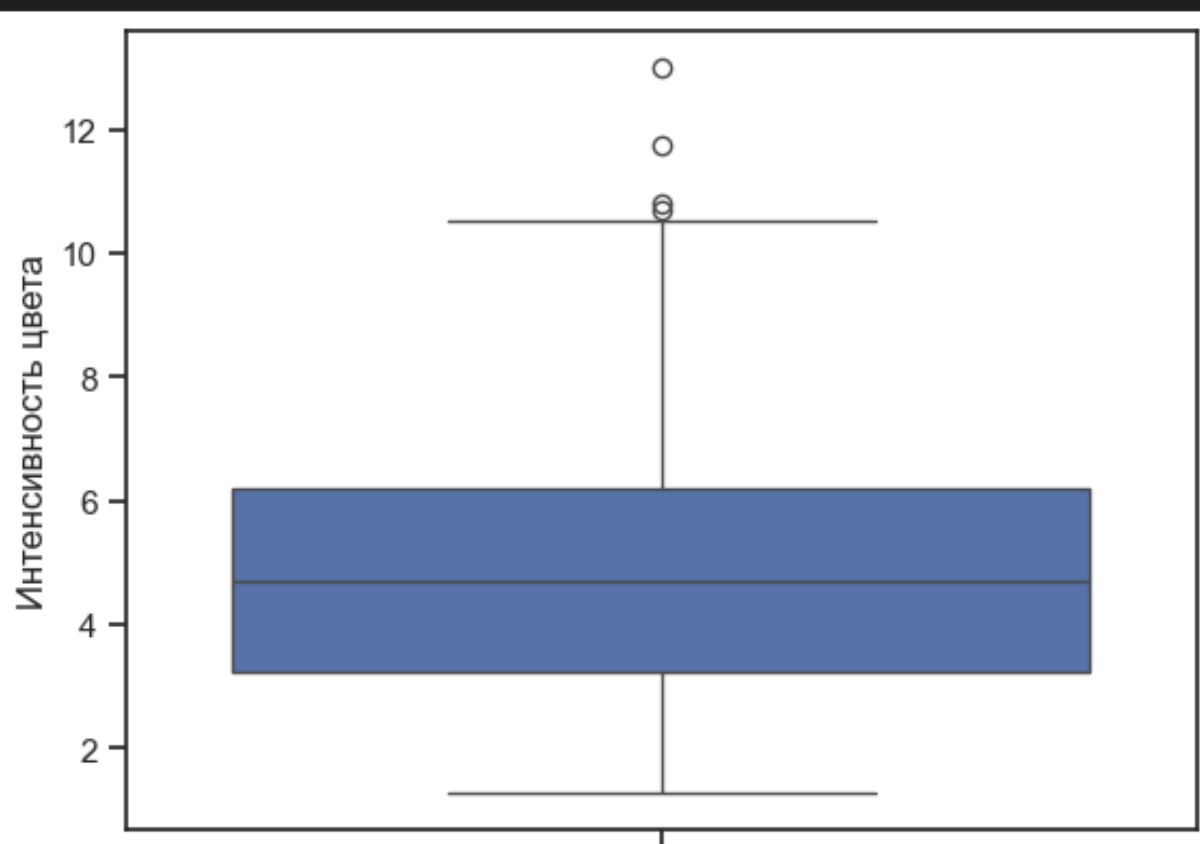
```
sns.boxplot(x=data['Интенсивность цвета'])
```

<Axes: xlabel='Интенсивность цвета'>



```
sns.boxplot(y=data['Интенсивность цвета'])
```

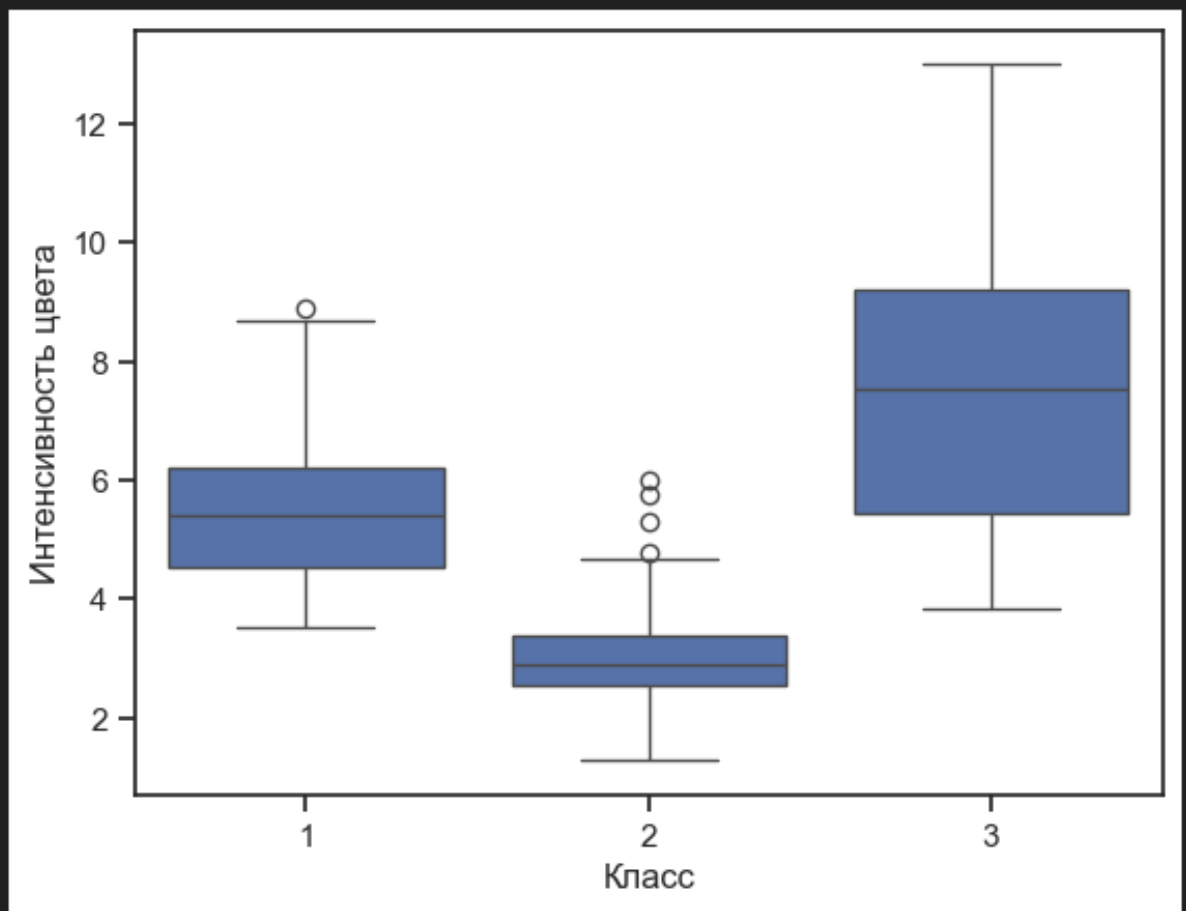
```
<Axes: ylabel='Интенсивность цвета'>
```



```
sns.boxplot(x='Класс', y='Интенсивность цвета', data=data)
```

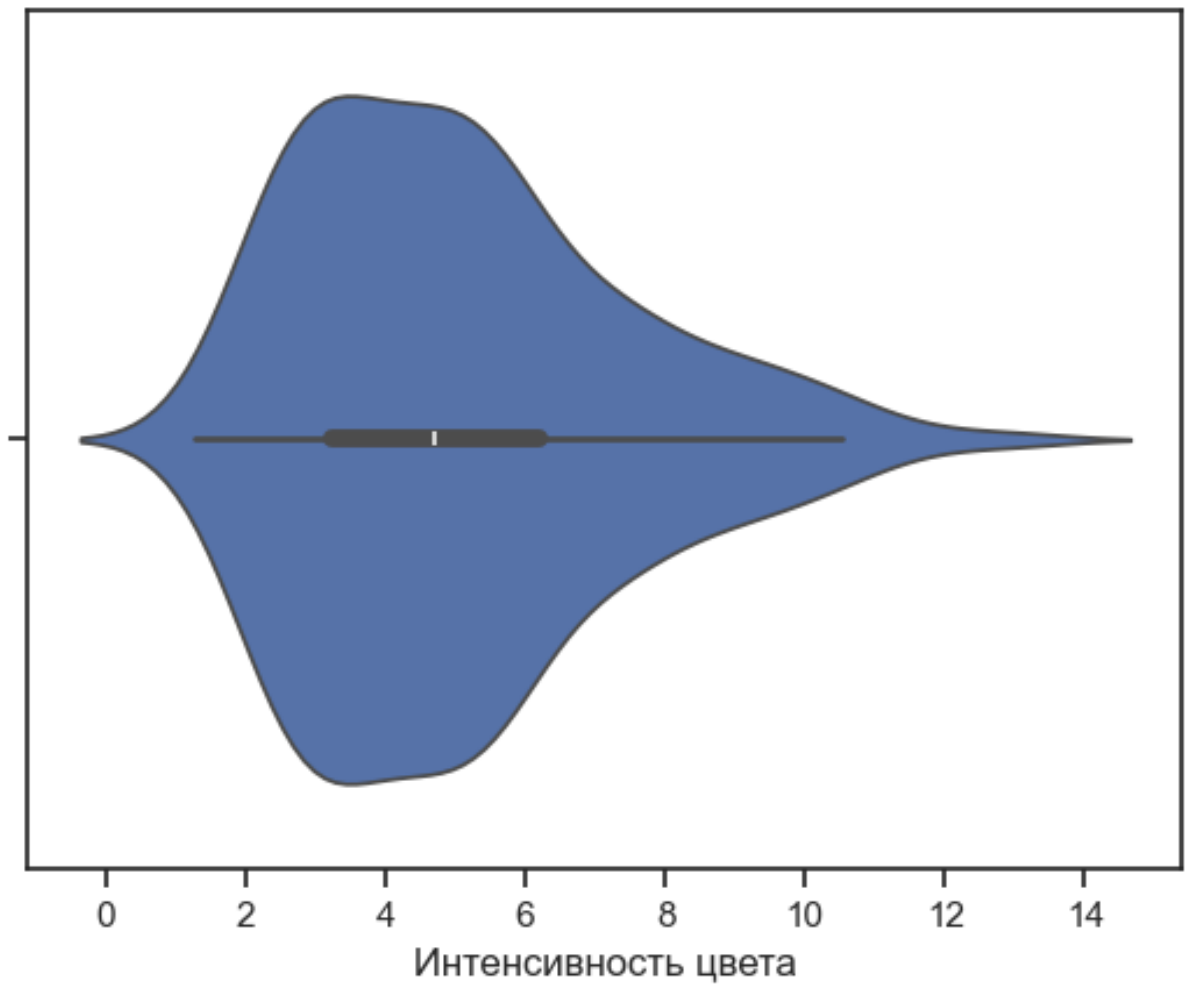
2]

```
<Axes: xlabel='Класс', ylabel='Интенсивность цвета'>
```



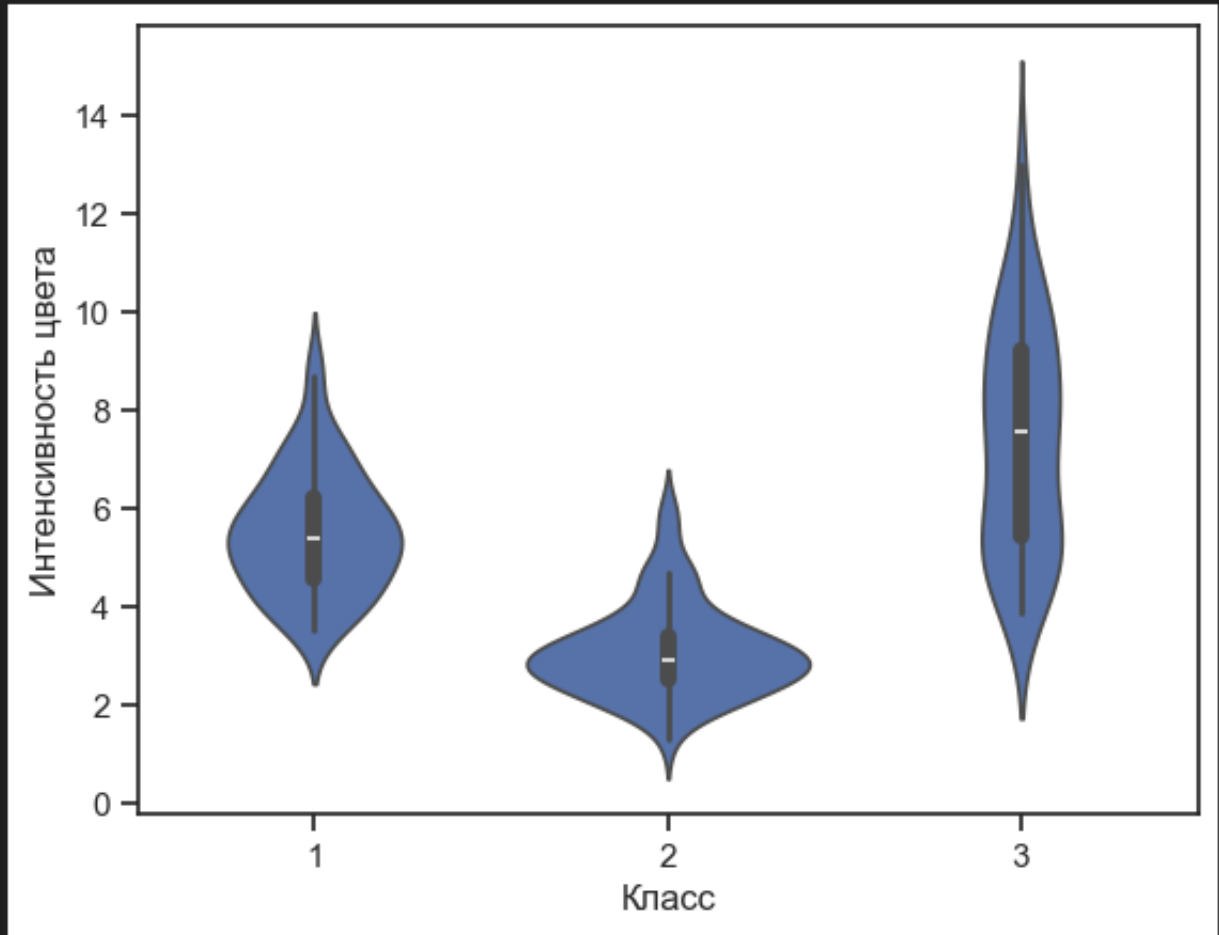
```
sns.violinplot(x=data['Интенсивность цвета'])
```

<Axes: xlabel='Интенсивность цвета'>




```
sns.violinplot(x='Класс', y='Интенсивность цвета', data=data)
```

```
<Axes: xlabel='Класс', ylabel='Интенсивность цвета'>
```



```
data.corr()
```

Python

	Класс	Алкоголь	Яблочная кислота	Пепел	Щелочность золы	Магний	Всего фенолов	Флаваноиды	Нефлаваноидные фенолы	Проантоцианы	Интенсивность цвета	Оттенок
Класс	1.000000	-0.328222	0.437776	-0.049643	0.517859	-0.209179	-0.719163	-0.847498	0.489109	-0.499130	0.265668	-0.617369
Алкоголь	-0.328222	1.000000	0.094397	0.211545	-0.310235	0.270798	0.289101	0.236815	-0.155929	0.136698	0.546364	-0.071747
Яблочная кислота	0.437776	0.094397	1.000000	0.164045	0.288500	-0.054575	-0.335167	-0.411007	0.292977	-0.220746	0.248985	-0.368710
Пепел	-0.049643	0.211545	0.164045	1.000000	0.443367	0.286587	0.128980	0.115077	0.186230	0.009652	0.258887	-0.074667
Щелочность золы	0.517859	-0.310235	0.288500	0.443367	1.000000	-0.083333	-0.321113	-0.351370	0.361922	-0.197327	0.018732	-0.276769
Магний	-0.209179	0.270798	-0.054575	0.286587	-0.083333	1.000000	0.214401	0.195784	-0.256294	0.236441	0.199950	0.066004
Всего фенолов	-0.719163	0.289101	-0.335167	0.128980	-0.321113	0.214401	1.000000	0.864564	-0.449935	0.612413	-0.055136	0.699949
Флаваноиды	-0.847498	0.236815	-0.411007	0.115077	-0.351370	0.195784	0.864564	1.000000	-0.537900	0.652692	-0.172379	0.787194
Нефлаваноидные фенолы	0.489109	-0.155929	0.292977	0.186230	0.361922	-0.256294	-0.449935	-0.537900	1.000000	-0.365845	0.139057	-0.503270
Проантоцианы	-0.499130	0.136698	-0.220746	0.009652	-0.197327	0.236441	0.612413	0.652692	-0.365845	1.000000	-0.025250	0.519067
Интенсивность цвета	0.265668	0.546364	0.248985	0.258887	0.018732	0.199950	-0.055136	-0.172379	0.139057	-0.025250	1.000000	-0.428815
Оттенок	-0.617369	-0.071747	-0.561296	-0.074667	-0.273955	0.055398	0.433681	0.543479	-0.262640	0.295544	-0.521813	1.000000
OD280/OD315 разбавленных вин	-0.788230	0.072343	-0.368710	0.003911	-0.276769	0.066004	0.699949	0.787194	-0.503270	0.519067	-0.428815	0.519067
Пролин	-0.633717	0.643720	-0.192011	0.223626	-0.440597	0.393351	0.498115	0.494193	-0.311385	0.330417	0.316100	0.498115

+ Code

+ Markdown

```
sns.heatmap(data.corr(), annot=True, fmt='.1f')
```

Python

<Axes: >

