**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)**

Курс «Технологии машинного обучения»

Отчет по лабораторной работе №1

«Разведочный анализ данных. Исследование и визуализация данных.»

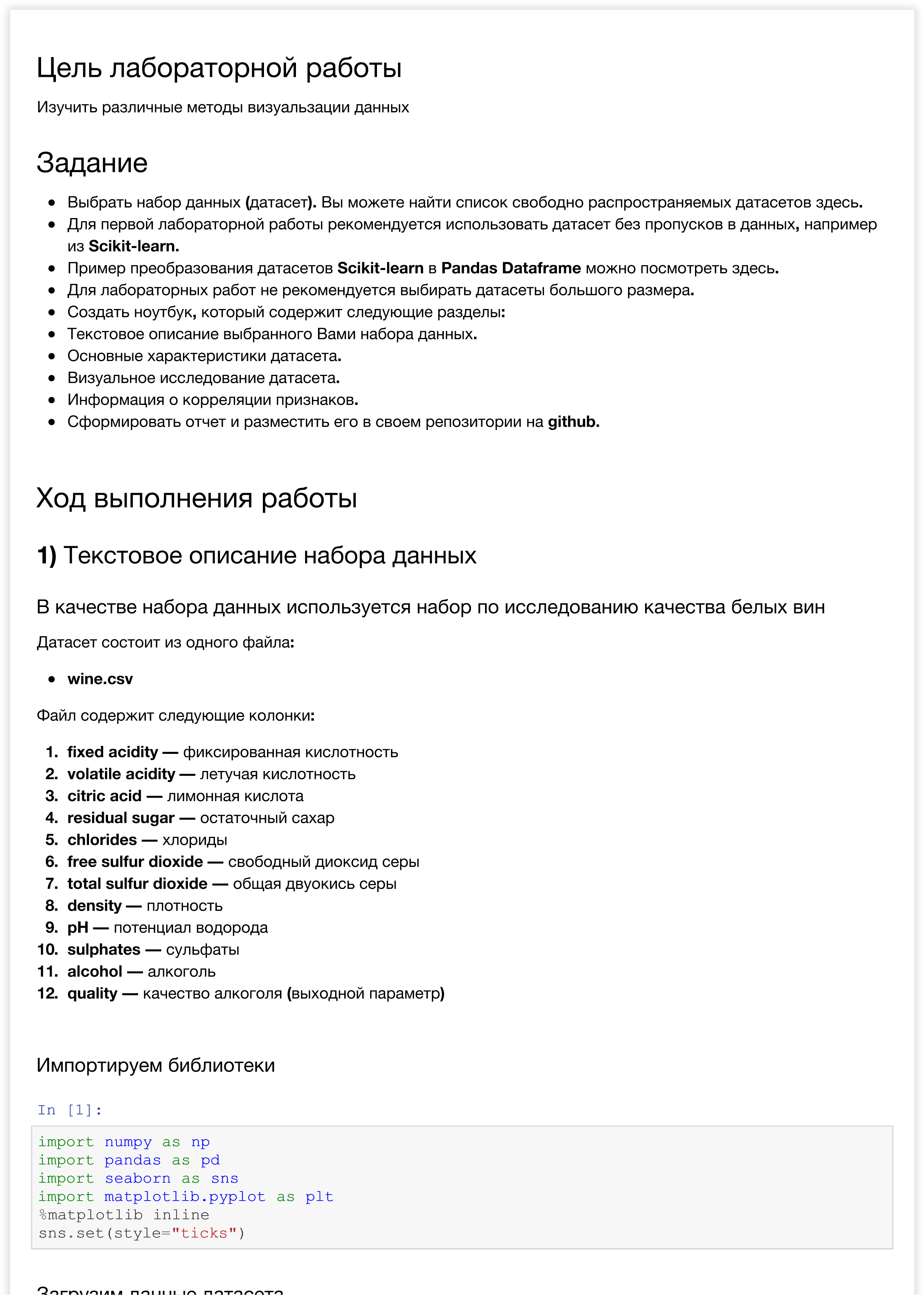
Выполнил:

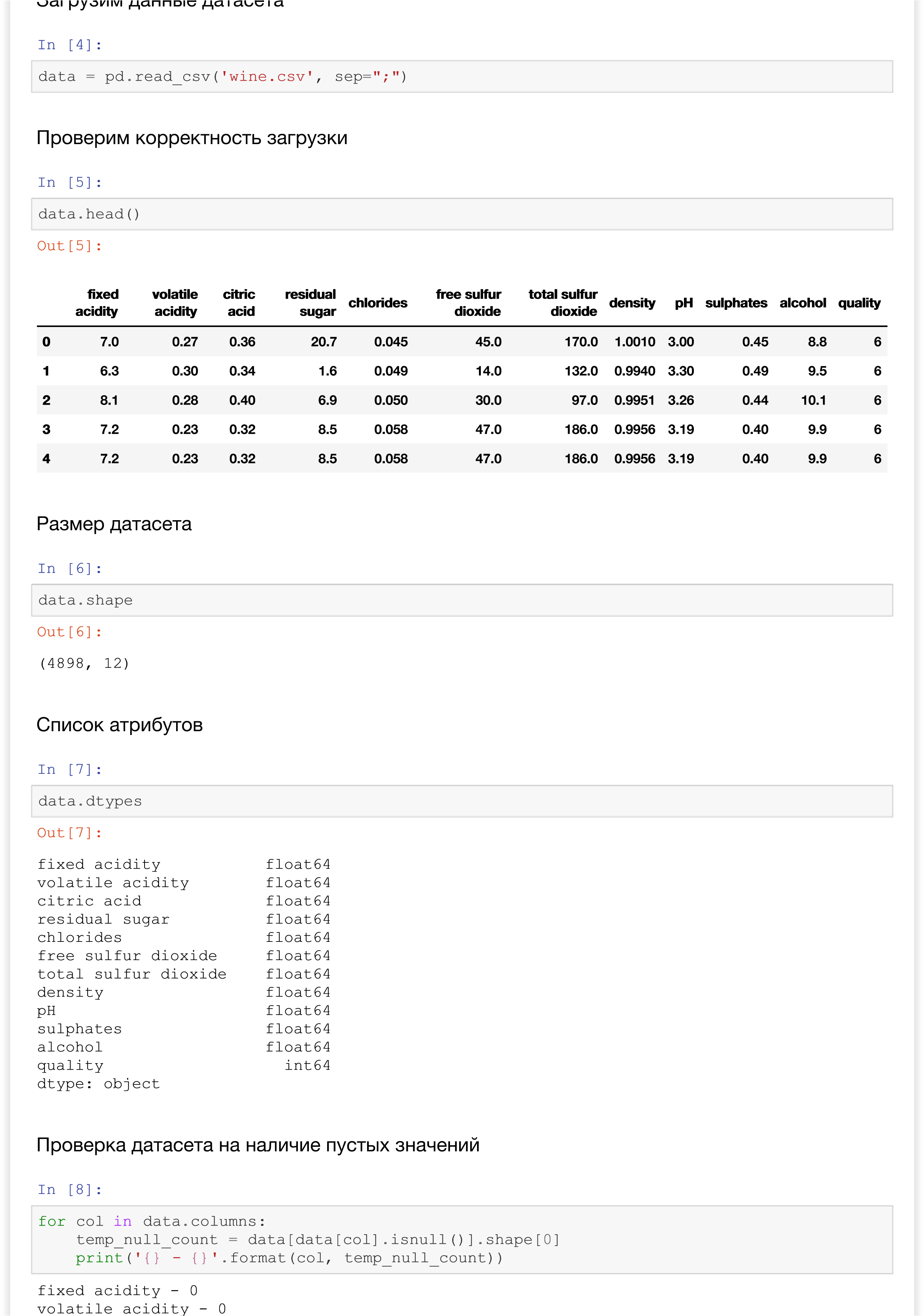
ИУ5Ц-82Б

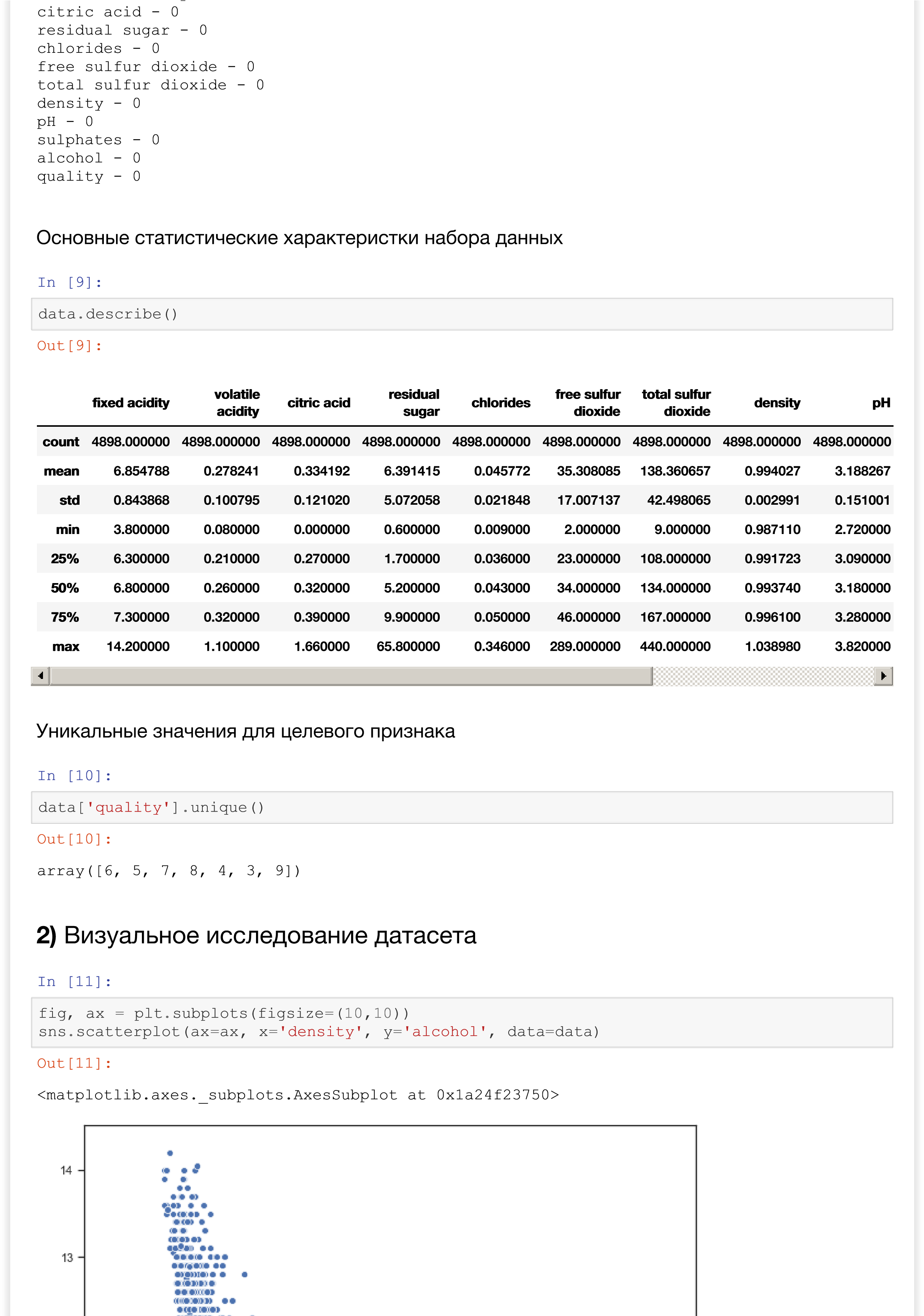
Гусев С.Р.

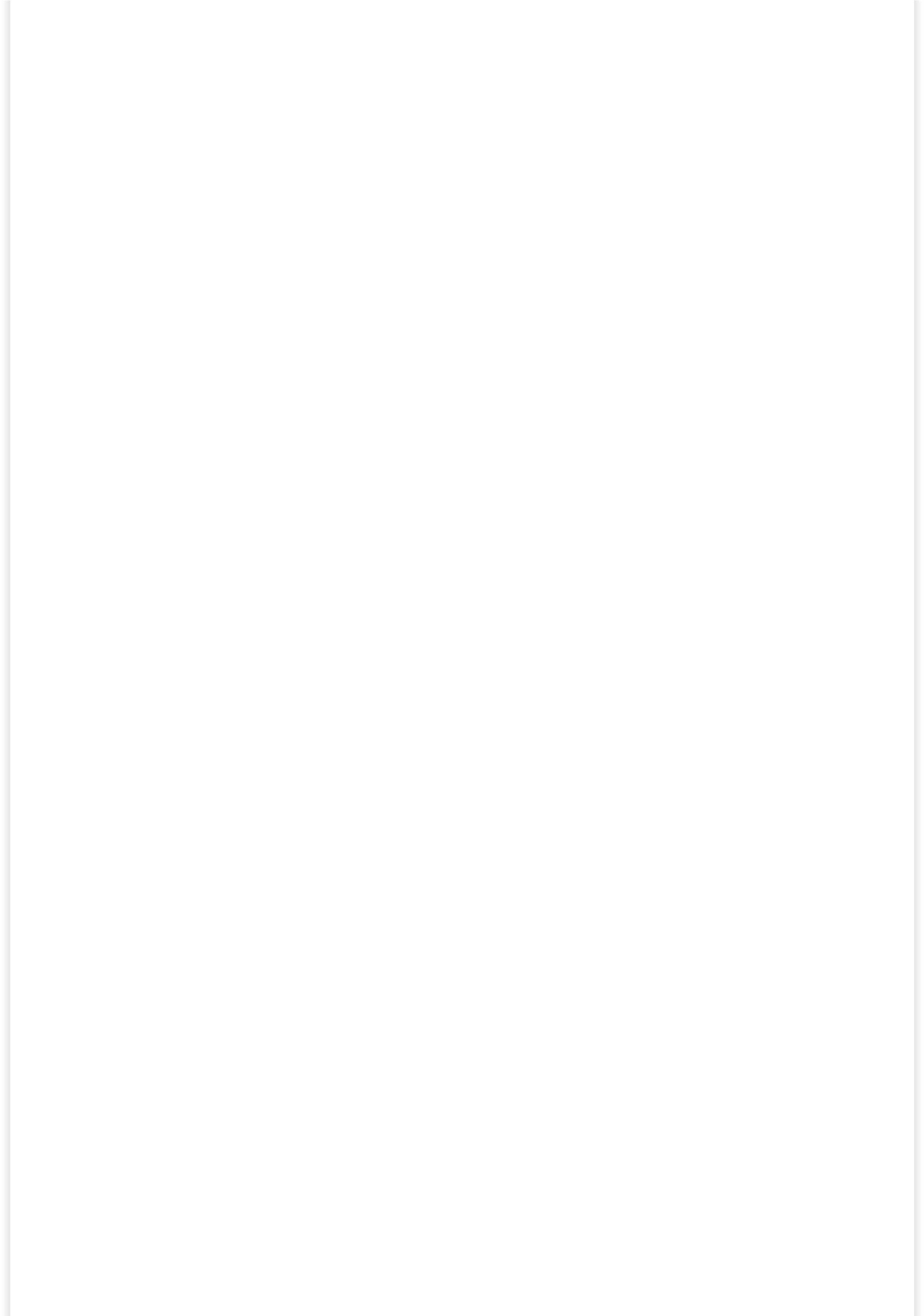
Преподаватель:

Гапанюк Ю.Е









Можно

заметить

**,**

что

между

атрибутами

**density**

и

**alcohol**

пристутствует

что

**-**

то

похожее

на

линейную

зависимость

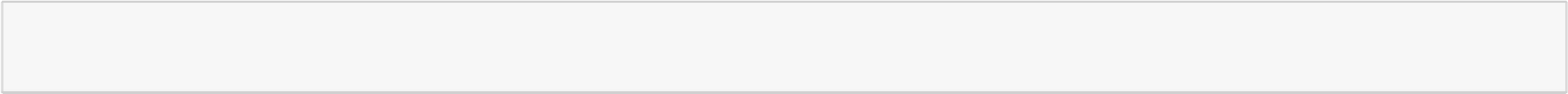
Введем

в

зависимость

целевой

признак



In [12]:

fig

,

ax

=

plt

.

subplots

(

figsize

=

(

8

,

8

))

sns

.

scatterplot

(

ax

=

ax

,

x

=

'density'

,

y

=

'alcohol'

,

data

=

data

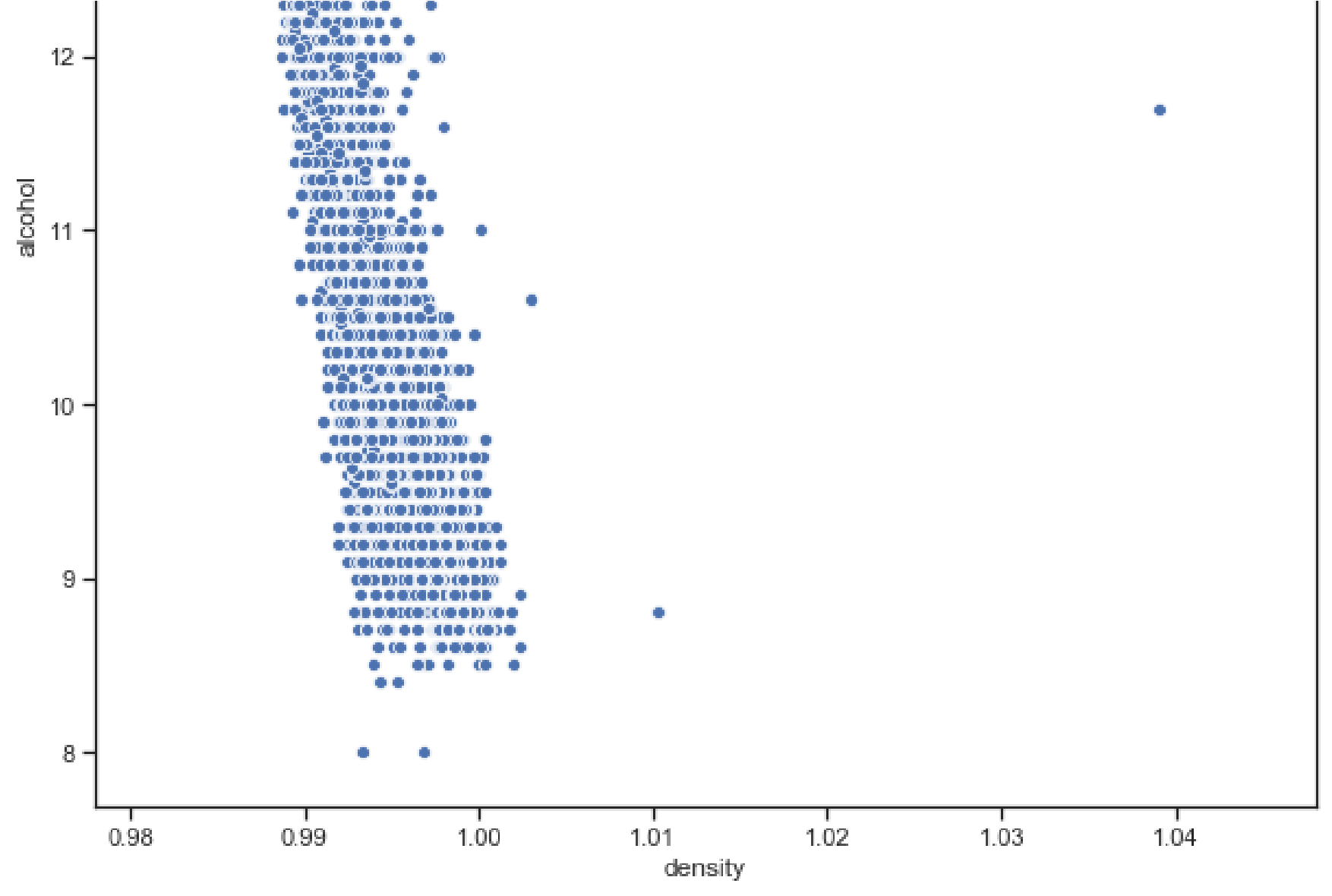
,

hue

=

'quality'

)

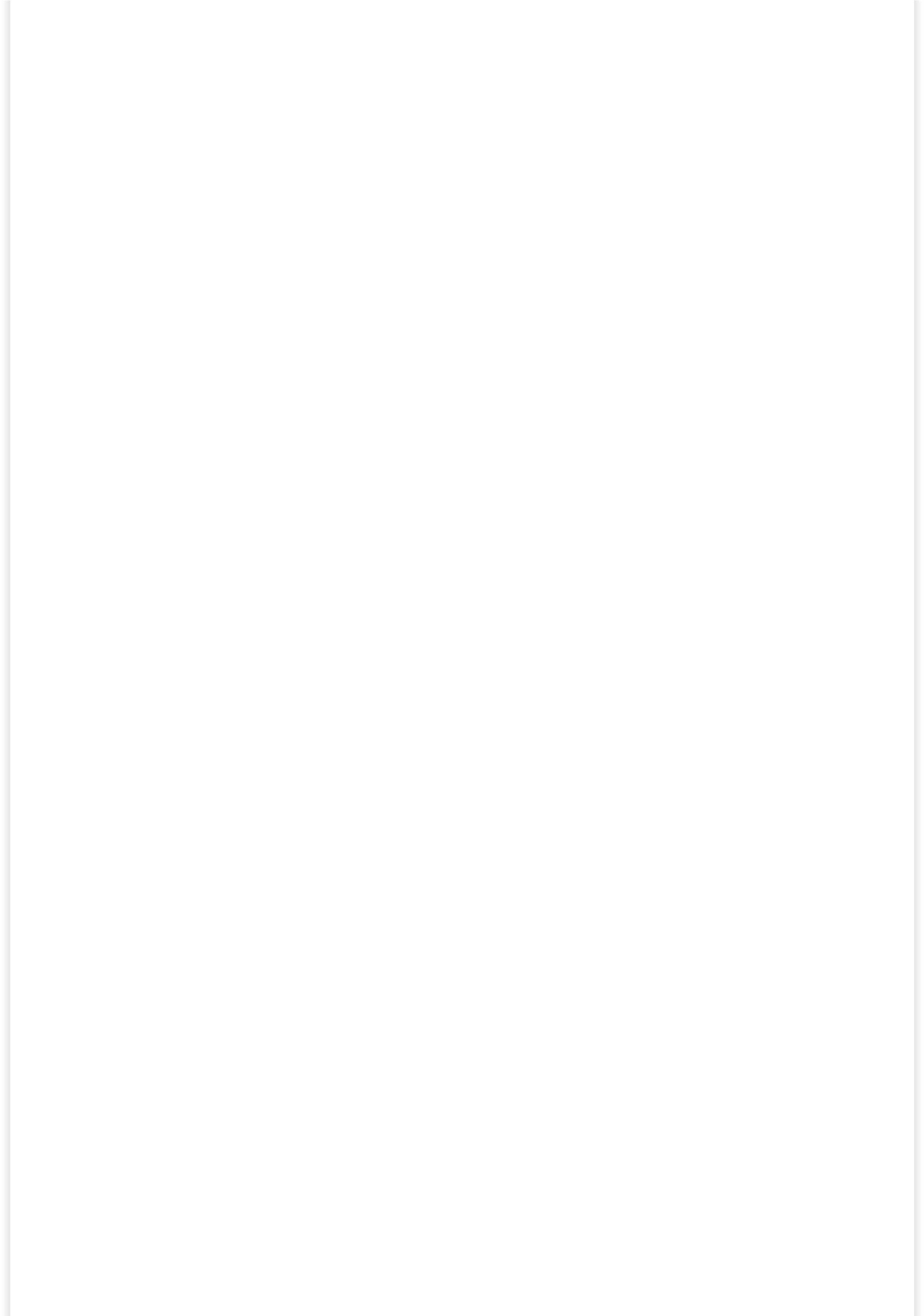


Out[12]:

<

matplotlib.axes.\_subplots.AxesSubplot at

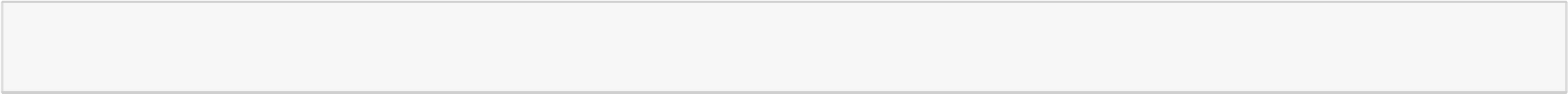
0x1a25d25090>



Распределение

качества

алкоголя



In [14]:

fig

,

ax

=

plt

.

subplots

(

figsize

=

(

13

,

13

))

sns

.

distplot

(

data

[

'quality'

]

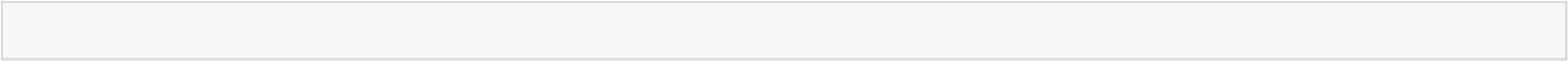
,

color

=

"g"

)



In [15]:

sns

.

jointplot

(

x

=

'density'

,

y

=

'alcohol'

,

data

=

data

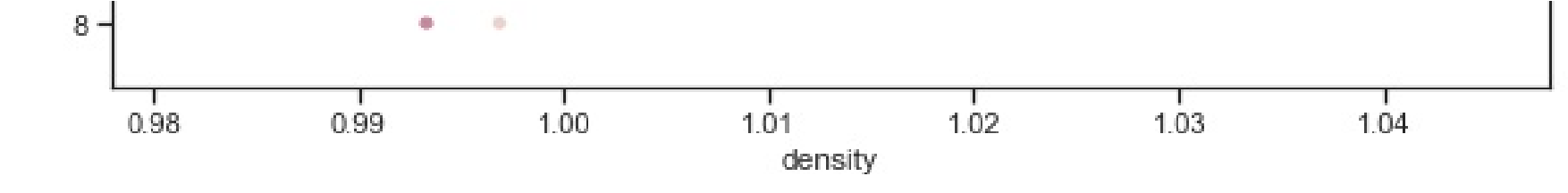
,

kind

=

"hex"

)

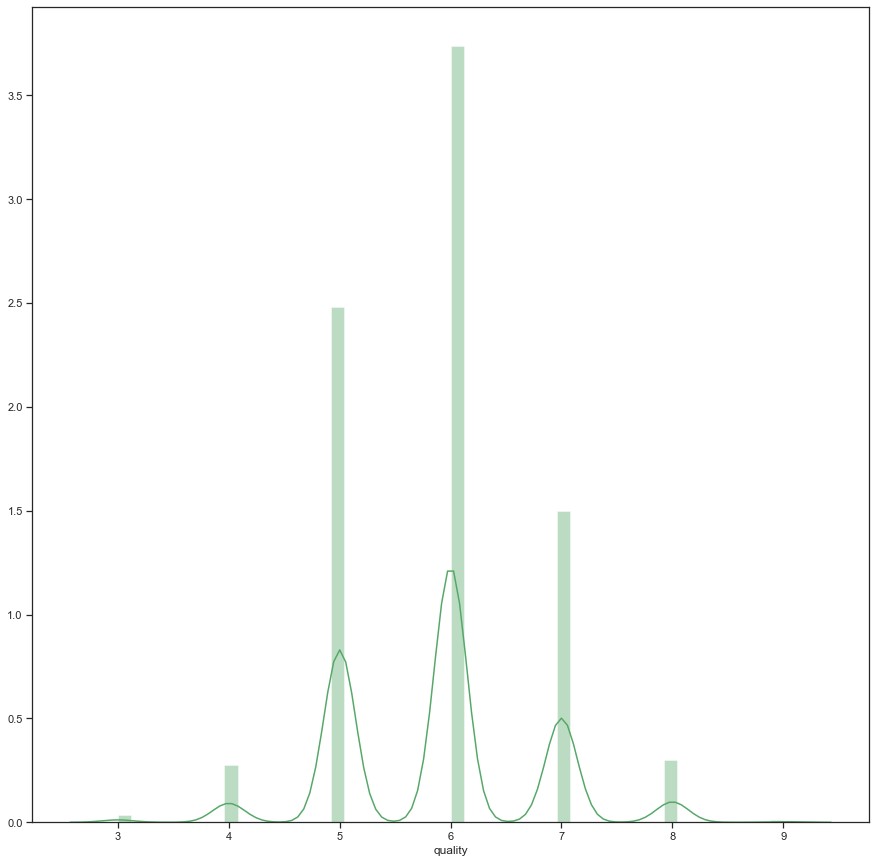


Out[14]:

<

0>

matplotlib.axes.\_subplots.AxesSubplot at 0x1a2635d6d

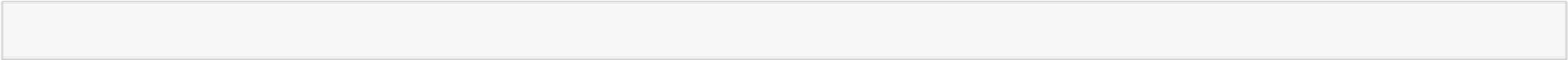
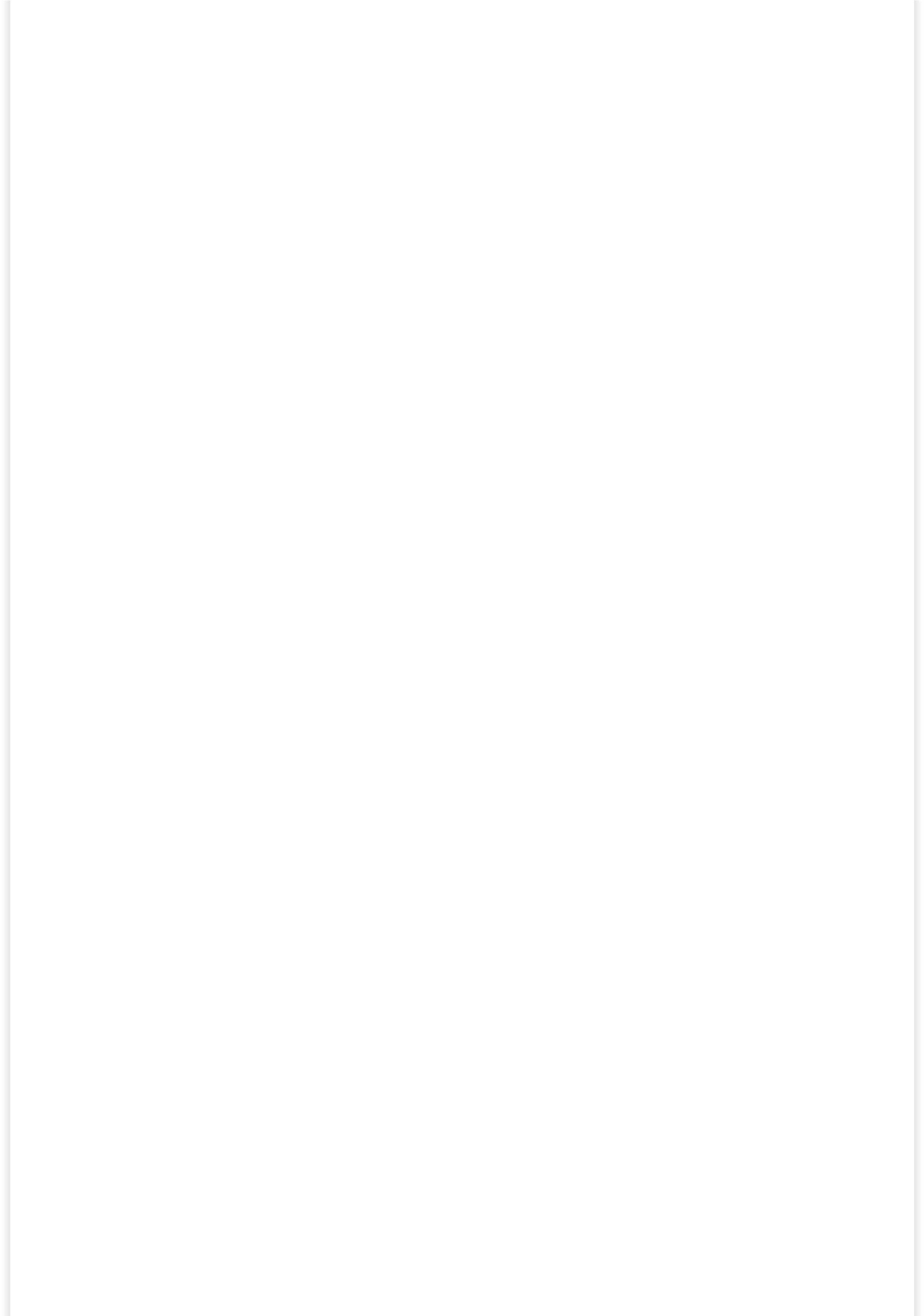


Out[15]:

<

seaborn.axisgrid.JointGrid at 0x1a26aee

650>



In [16]:

sns

.

jointplot

(

x

=

'density'

,

y

=

'alcohol'

,

data

=

data

,

kind

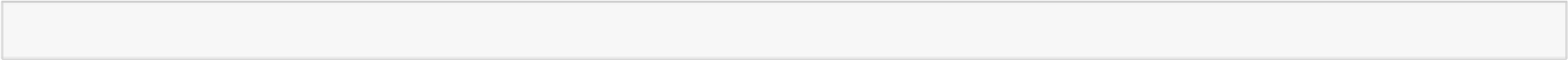
=

"kde"

)

Парные

диаграммы



In [17]:

sns

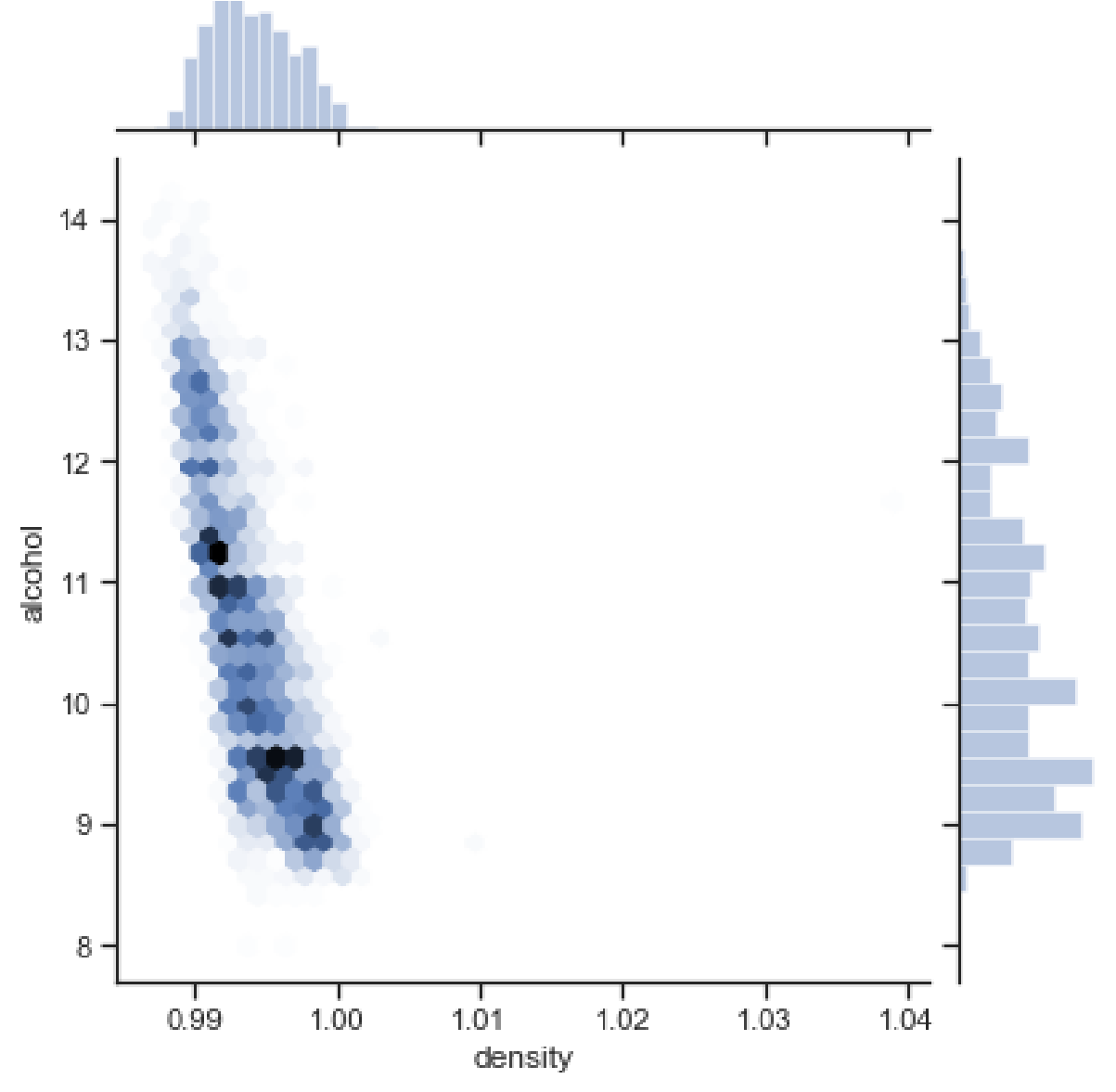
.

pairplot

(

data

)

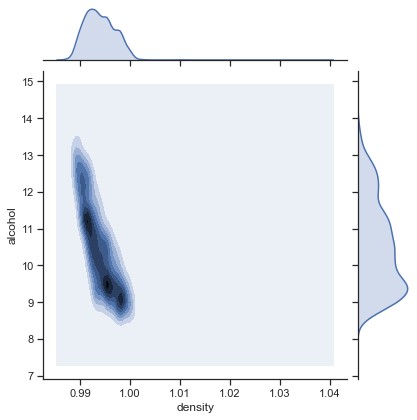


Out[16]:

<

seaborn.axisgrid.JointGrid at

0x1a25e87890>

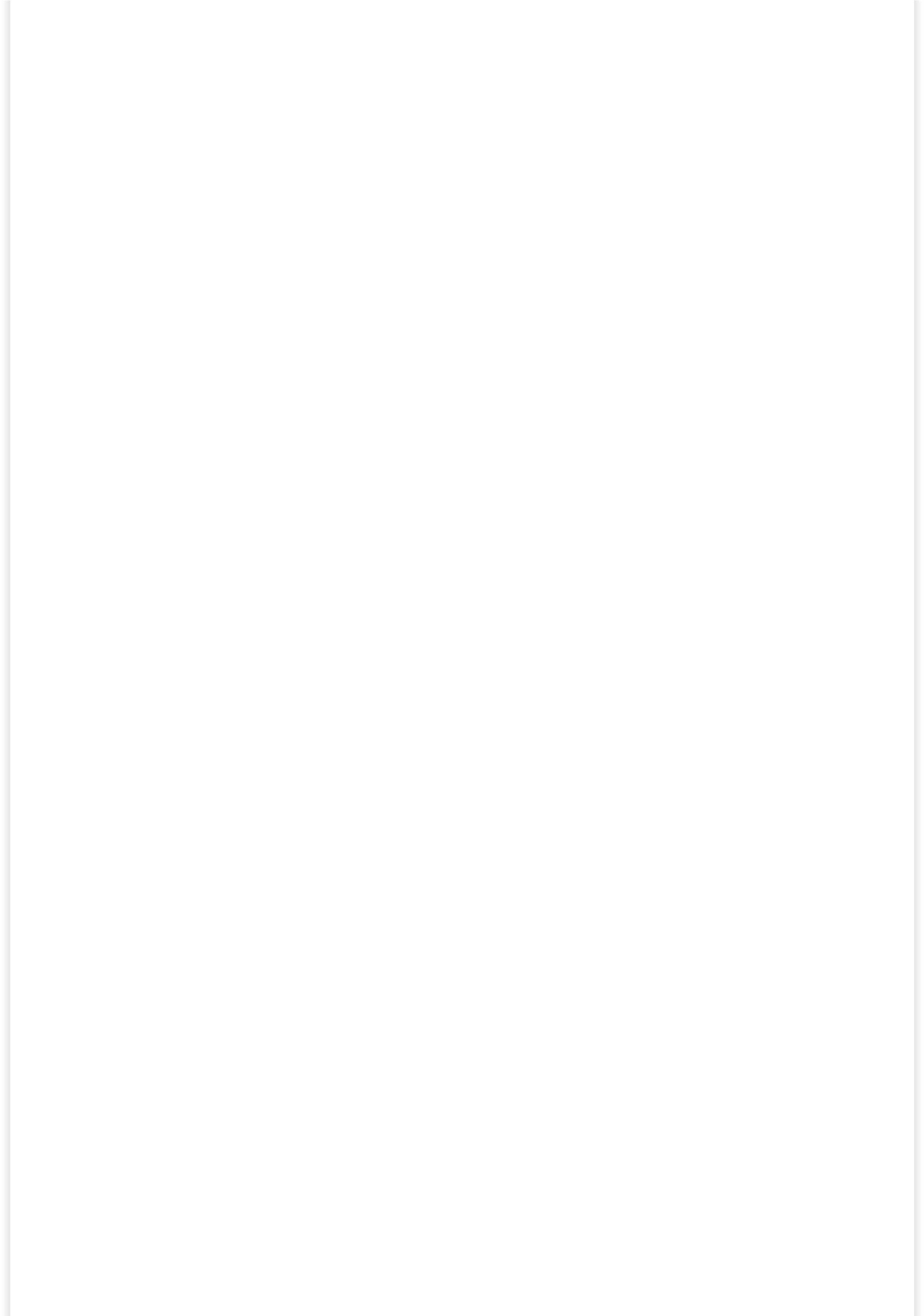
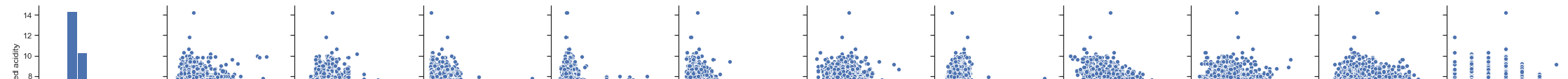


Out[17]:

<

seaborn.axisgrid.PairGrid at 0x1a2740eb

10>



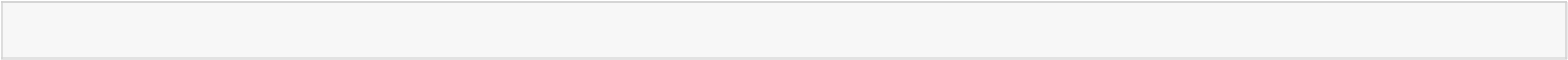
Группирование

по

значению

целевого

признака



In [18]:

sns

.

pairplot

(

data

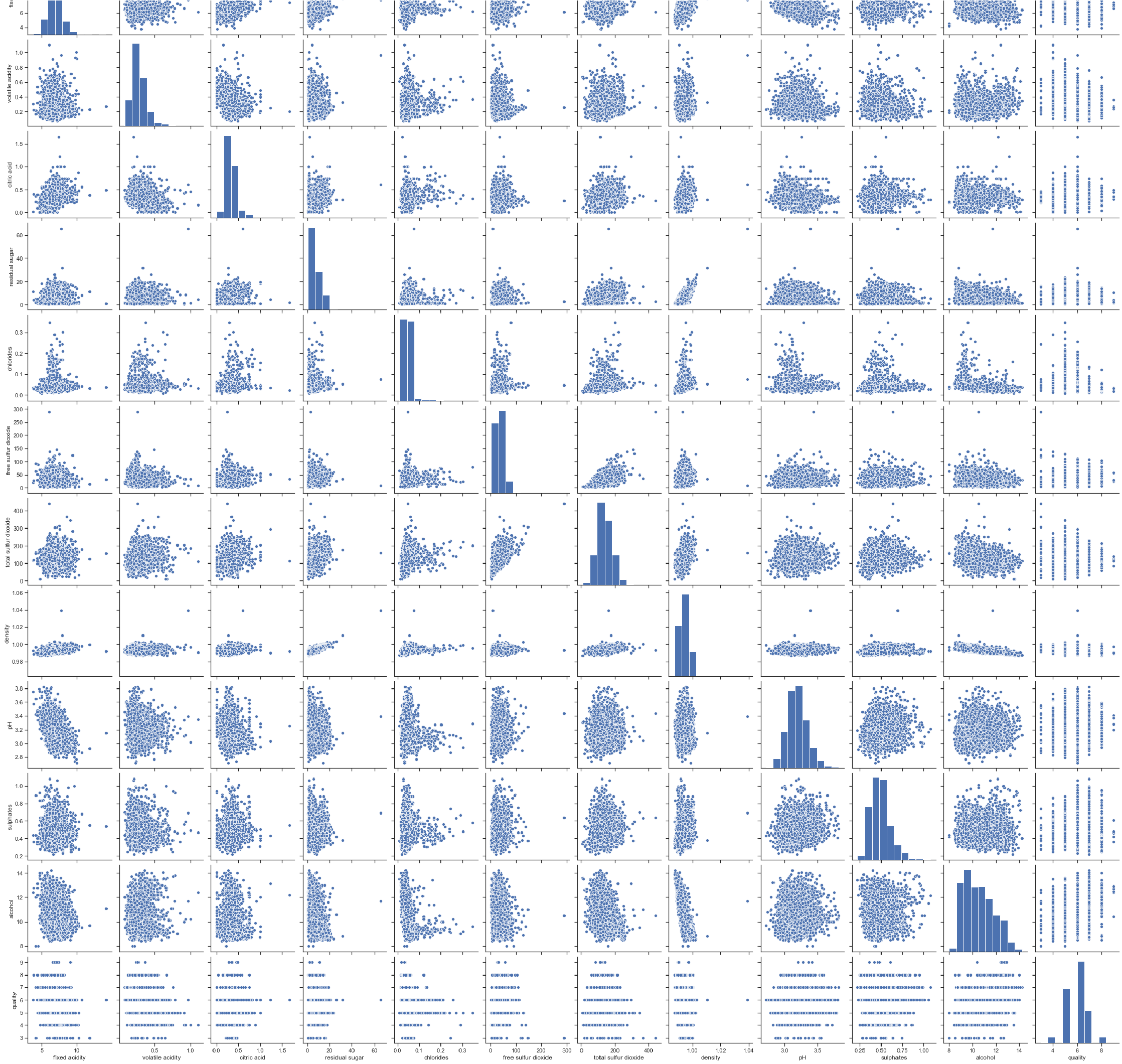
,

hue

=

"quality"

)

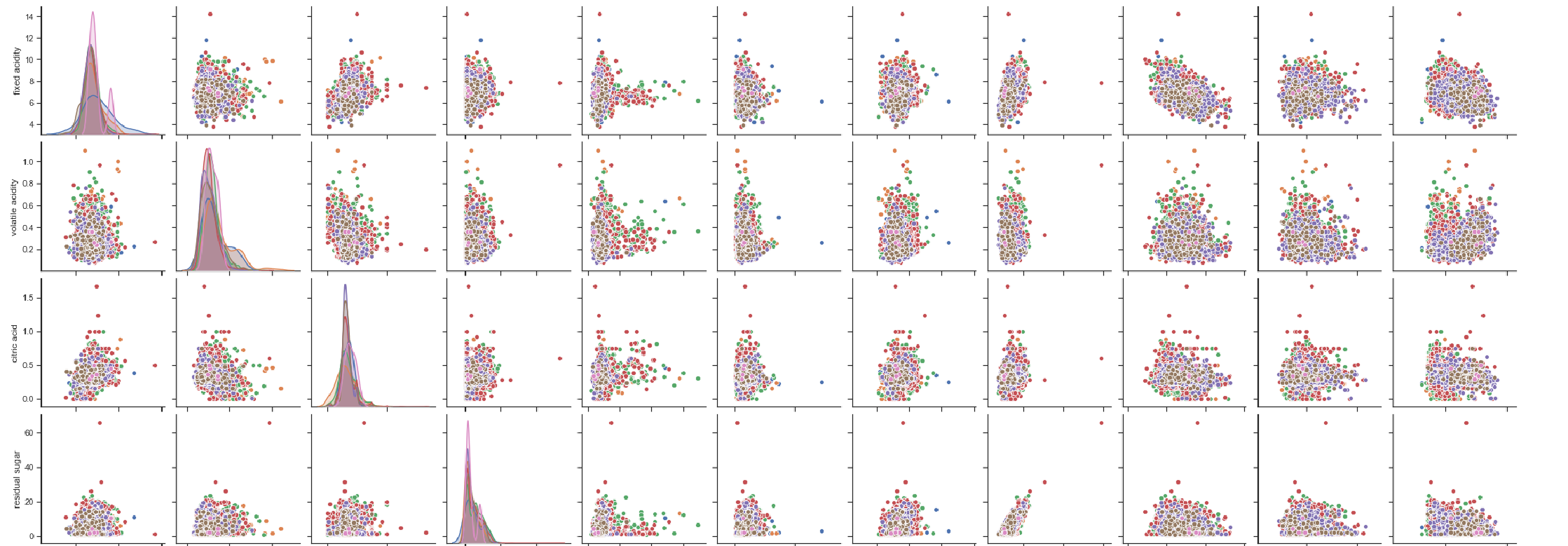


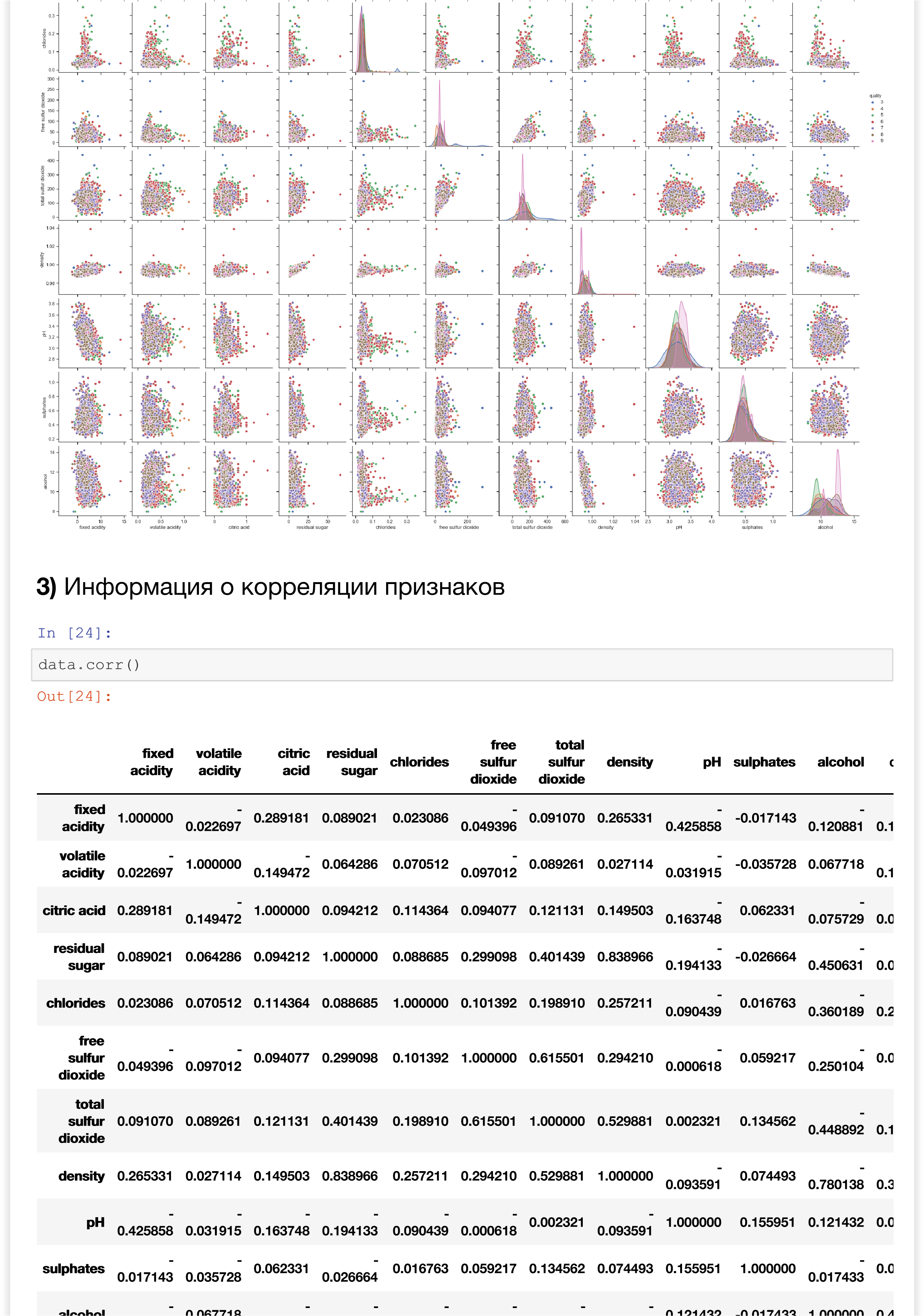
Out[18]:

<

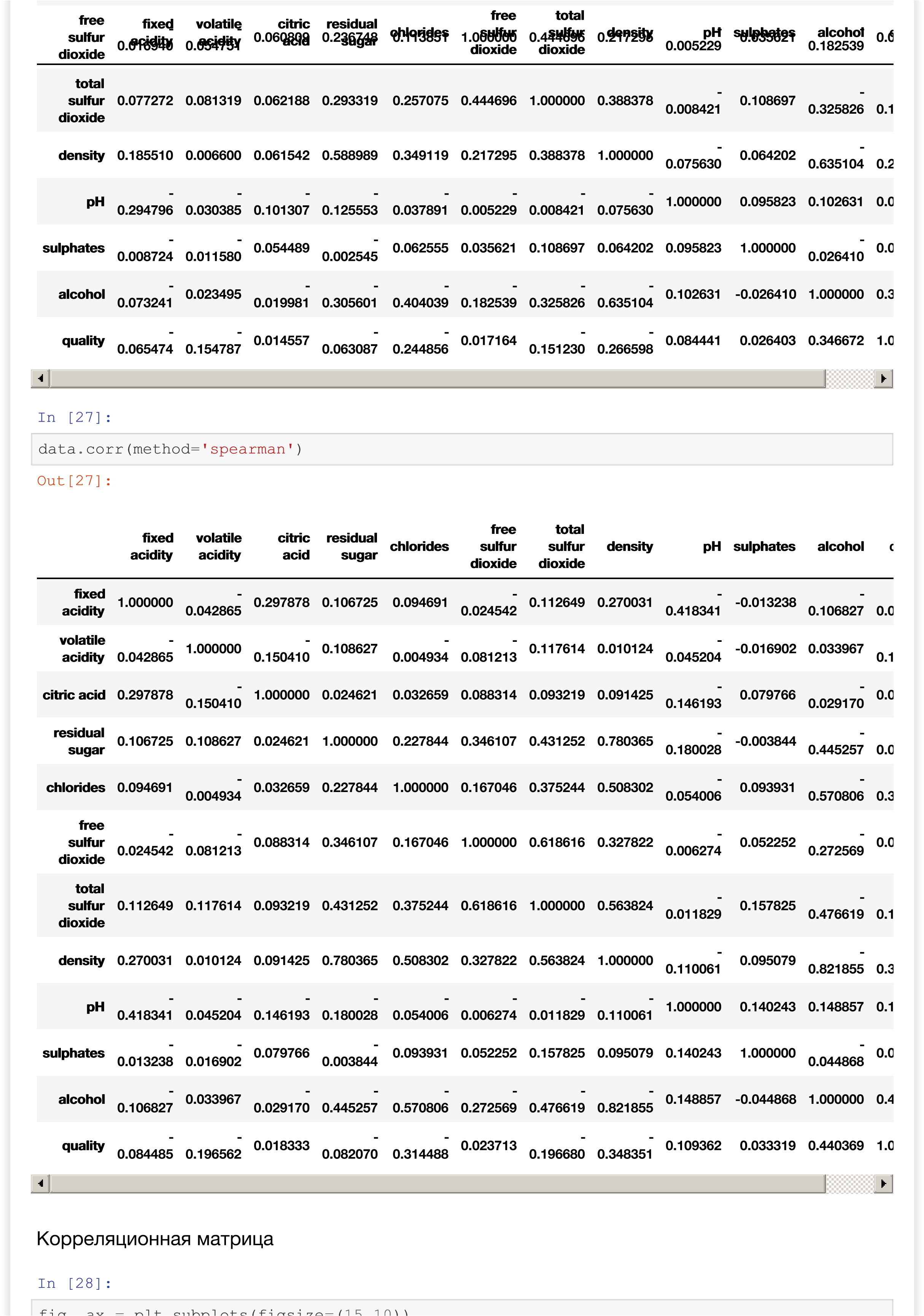
seaborn.axisgrid.PairGrid at 0x1a

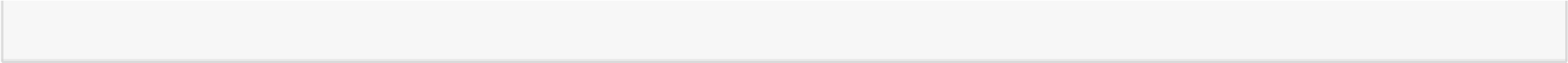
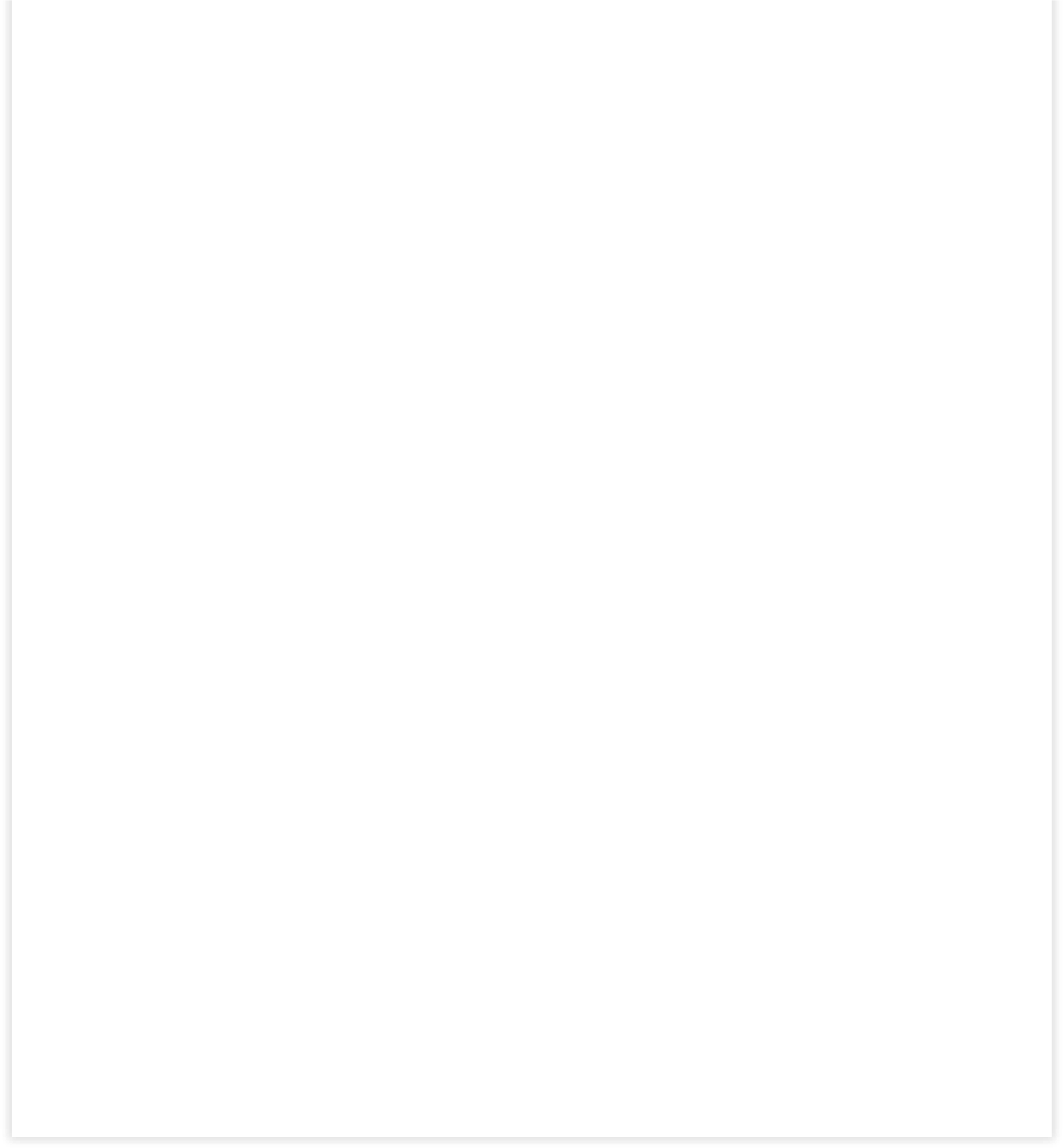
2c014b10>











fig

,

ax

=

plt

.

subplots

(

figsize

=

(

15

,

10

))

sns

.

heatmap

(

data

.

corr

()

,

annot

=

True

,

fmt

=

'.2f'

,

cmap

=

'GnBu'

)

Выводы

о

коррелирующих

признаках

**1**

**.**

Коэффициенты

корреляции

в

данном

наборе

низкие

**.**

**2**

**.**

Входные

параметры

влияют

на

качество

алкоголя

**,**

так

как

они

составляют

его

химический

состав

**.**

**3**

**.**

**'alcohol'**

и

**'density'**

лучше

всего

коррелируют

с

целевым

признаком

**,**

однако

они

очень

сильно

коррелируют

друг

с

другом

**(|0.78|).**

**4**

**.**

**'free sulfur'**

и

**'total sulfur'**

довольно

неплохо

коррелируют

друг

с

другом

**(|0.62|),**

что

логично

**,**

так

как

общий

дикосид

серы

является

сумма

связной

и

свободной

серы

**.**

У

них

прослеживается

явная

заивисмость

**.**

Out[28]:

<

matplotlib.axes.\_subplots.AxesSubplot at 0x1a2df

1a590>

