Рубежный контроль №1

Группа: ИУ5Ц-82Б

Номер варианта 26

Студент: Бабаян Артур Ашотович

Задание:

Необходимо подготовить отчет по рубежному контролю и разместить его в Вашем репозитории. Вы можете использовать титульный лист, или в начале ноутбука в текстовой ячейке указать Ваши Ф.И.О. и группу.

Импорт библиотек

```
import sys
sys.path
import pandas as pd
import numpy as np
np.seterr(divide='ignore', invalid='ignore')
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

Загрузка данных

```
print(sys.version)
```

```
3.10.0 (tags/v3.10.0:b494f59, Oct  4 2021, 19:00:18) [MSC v.1929 64
bit (AMD64)]
```

```
from sklearn.datasets import load_wine
```

```
data = load_wine()
```

Основные характеристики датасета

```
# Вывод фактических данных.
data
```

```
{'data': array([[1.423e+01, 1.710e+00, 2.430e+00, ..., 1.040e+00,
3.920e+00,
         1.065e+03],
        [1.320e+01, 1.780e+00, 2.140e+00, ..., 1.050e+00, 3.400e+00,
         1.050e+03],
        [1.316e+01, 2.360e+00, 2.670e+00, ..., 1.030e+00, 3.170e+00,
         1.185e+03],
        ...,
        [1.327e+01, 4.280e+00, 2.260e+00, ..., 5.900e-01, 1.560e+00,
         8.350e+02],
```

```
        [1.317e+01, 2.590e+00, 2.370e+00, ..., 6.000e-01, 1.620e+00,
         8.400e+02],
        [1.413e+01, 4.100e+00, 2.740e+00, ..., 6.100e-01, 1.600e+00,
         5.600e+02]]),
 'target': array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1,
1,
        1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1,
        1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1,
        1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2,
2,
        2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
2,
        2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
2,
        2, 2]),
 'frame': None,
 'target_names': array(['class_0', 'class_1', 'class_2'],
dtype='<U7'),
 'DESCR': '.. _wine_dataset:\n\nWine recognition dataset\
n------------------------\n\n**Data Set Characteristics:**\n\
n    :Number of Instances: 178\n    :Number of Attributes: 13 numeric,
predictive attributes and the class\n    :Attribute Information:\n \t\
t- Alcohol\n \t\t- Malic acid\n \t\t- Ash\n\t\t- Alcalinity of ash  \n
\t\t- Magnesium\n\t\t- Total phenols\n \t\t- Flavanoids\n \t\t-
Nonflavanoid phenols\n \t\t- Proanthocyanins\n\t\t- Color intensity\
n \t\t- Hue\n \t\t- OD280/OD315 of diluted wines\n \t\t- Proline\n\n
- class:\n            - class_0\n            - class_1\n            -
class_2\n\t\t\n    :Summary Statistics:\n    \n
============================= ==== ===== ======= =====\n
Min   Max   Mean     SD\n    ============================= ==== =====
======= =====\n    Alcohol:                          11.0  14.8    13.0
0.8\n    Malic Acid:                 0.74  5.80    2.34  1.12\n
Ash:                          1.36  3.23    2.36  0.27\n    Alcalinity
of Ash:            10.6  30.0    19.5   3.3\n    Magnesium:
70.0 162.0    99.7  14.3\n    Total Phenols:                0.98  3.88
2.29  0.63\n    Flavanoids:                   0.34  5.08    2.03
1.00\n    Nonflavanoid Phenols:         0.13  0.66    0.36  0.12\n
Proanthocyanins:              0.41  3.58    1.59  0.57\n    Colour
Intensity:             1.3  13.0     5.1   2.3\n    Hue:
0.48  1.71    0.96  0.23\n    OD280/OD315 of diluted wines: 1.27  4.00
2.61  0.71\n    Proline:                       278  1680     746
315\n    ============================= ==== ===== ======= =====\n\n
:Missing Attribute Values: None\n    :Class Distribution: class_0
(59), class_1 (71), class_2 (48)\n    :Creator: R.A. Fisher\
```

```
n    :Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov)\
n    :Date: July, 1988\n\nThis is a copy of UCI ML Wine recognition
datasets.\nhttps://archive.ics.uci.edu/ml/machine-learning-databases/
wine/wine.data\n\nThe data is the results of a chemical analysis of
wines grown in the same\nregion in Italy by three different
cultivators. There are thirteen different\nmeasurements taken for
different constituents found in the three types of\nwine.\n\nOriginal
Owners: \n\nForina, M. et al, PARVUS - \nAn Extendible Package for
Data Exploration, Classification and Correlation. \nInstitute of
Pharmaceutical and Food Analysis and Technologies,\nVia Brigata
Salerno, 16147 Genoa, Italy.\n\nCitation:\n\nLichman, M. (2013). UCI
Machine Learning Repository\n[https://archive.ics.uci.edu/ml]. Irvine,
CA: University of California,\nSchool of Information and Computer
Science. \n\n.. topic:: References\n\n  (1) S. Aeberhard, D. Coomans
and O. de Vel, \n  Comparison of Classifiers in High Dimensional
Settings, \n  Tech. Rep. no. 92-02, (1992), Dept. of Computer Science
and Dept. of  \n  Mathematics and Statistics, James Cook University of
North Queensland. \n  (Also submitted to Technometrics). \n\n  The
data was used with many others for comparing various \n  classifiers.
The classes are separable, though only RDA \n  has achieved 100%
correct classification. \n  (RDA : 100%, QDA 99.4%, LDA 98.9%, 1NN
96.1% (z-transformed data)) \n  (All results using the leave-one-out
technique) \n\n  (2) S. Aeberhard, D. Coomans and O. de Vel, \n  "THE
CLASSIFICATION PERFORMANCE OF RDA" \n  Tech. Rep. no. 92-01, (1992),
Dept. of Computer Science and Dept. of \n  Mathematics and Statistics,
James Cook University of North Queensland. \n  (Also submitted to
Journal of Chemometrics).\n',
 'feature_names': ['alcohol',
  'malic_acid',
  'ash',
  'alcalinity_of_ash',
  'magnesium',
  'total_phenols',
  'flavanoids',
  'nonflavanoid_phenols',
  'proanthocyanins',
  'color_intensity',
  'hue',
  'od280/od315_of_diluted_wines',
  'proline']}
```

```
# Вывод независимых переменных.
data.data
```

```
array([[1.423e+01, 1.710e+00, 2.430e+00, ..., 1.040e+00, 3.920e+00,
        1.065e+03],
       [1.320e+01, 1.780e+00, 2.140e+00, ..., 1.050e+00, 3.400e+00,
        1.050e+03],
       [1.316e+01, 2.360e+00, 2.670e+00, ..., 1.030e+00, 3.170e+00,
        1.185e+03],
       ...,
```

```
       [1.327e+01, 4.280e+00, 2.260e+00, ..., 5.900e-01, 1.560e+00,
        8.350e+02],
       [1.317e+01, 2.590e+00, 2.370e+00, ..., 6.000e-01, 1.620e+00,
        8.400e+02],
       [1.413e+01, 4.100e+00, 2.740e+00, ..., 6.100e-01, 1.600e+00,
        5.600e+02]])
```

```python
# Вывод столбцов.
data.feature_names
```

```
['alcohol',
 'malic_acid',
 'ash',
 'alcalinity_of_ash',
 'magnesium',
 'total_phenols',
 'flavanoids',
 'nonflavanoid_phenols',
 'proanthocyanins',
 'color_intensity',
 'hue',
 'od280/od315_of_diluted_wines',
 'proline']
```

```python
# Вывод описания.
data.DESCR
```

```
'.. _wine_dataset:\n\nWine recognition dataset\
n------------------------\n\n**Data Set Characteristics:**\n\
n    :Number of Instances: 178\n    :Number of Attributes: 13 numeric,
predictive attributes and the class\n    :Attribute Information:\n \t\
t- Alcohol\n \t\t- Malic acid\n \t\t- Ash\n\t\t- Alcalinity of ash  \n
\t\t- Magnesium\n\t\t- Total phenols\n \t\t- Flavanoids\n \t\t-
Nonflavanoid phenols\n \t\t- Proanthocyanins\n\t\t- Color intensity\
n \t\t- Hue\n \t\t- OD280/OD315 of diluted wines\n \t\t- Proline\n\n
- class:\n            - class_0\n            - class_1\n            -
class_2\n\t\t\n    :Summary Statistics:\n    \n
============================= ==== ===== ======= =====\n
Min   Max   Mean     SD\n    ============================= ==== =====
======= =====\n    Alcohol:                          11.0  14.8    13.0
0.8\n    Malic Acid:                   0.74  5.80    2.34  1.12\n
Ash:                          1.36  3.23    2.36  0.27\n    Alcalinity
of Ash:            10.6  30.0    19.5   3.3\n    Magnesium:
70.0 162.0    99.7  14.3\n    Total Phenols:                0.98  3.88
2.29  0.63\n    Flavanoids:                   0.34  5.08    2.03
1.00\n    Nonflavanoid Phenols:         0.13  0.66    0.36  0.12\n
Proanthocyanins:              0.41  3.58    1.59  0.57\n    Colour
Intensity:              1.3  13.0     5.1   2.3\n    Hue:
0.48  1.71    0.96  0.23\n    OD280/OD315 of diluted wines: 1.27  4.00
2.61  0.71\n    Proline:                       278  1680     746
315\n    ============================= ==== ===== ======= =====\n\n
```

```
:Missing Attribute Values: None\n    :Class Distribution: class_0
(59), class_1 (71), class_2 (48)\n    :Creator: R.A. Fisher\
n    :Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov)\
n    :Date: July, 1988\n\nThis is a copy of UCI ML Wine recognition
datasets.\nhttps://archive.ics.uci.edu/ml/machine-learning-databases/
wine/wine.data\n\nThe data is the results of a chemical analysis of
wines grown in the same\nregion in Italy by three different
cultivators. There are thirteen different\nmeasurements taken for
different constituents found in the three types of\nwine.\n\nOriginal
Owners: \n\nForina, M. et al, PARVUS - \nAn Extendible Package for
Data Exploration, Classification and Correlation. \nInstitute of
Pharmaceutical and Food Analysis and Technologies,\nVia Brigata
Salerno, 16147 Genoa, Italy.\n\nCitation:\n\nLichman, M. (2013). UCI
Machine Learning Repository\n[https://archive.ics.uci.edu/ml]. Irvine,
CA: University of California,\nSchool of Information and Computer
Science. \n\n.. topic:: References\n\n  (1) S. Aeberhard, D. Coomans
and O. de Vel, \n  Comparison of Classifiers in High Dimensional
Settings, \n  Tech. Rep. no. 92-02, (1992), Dept. of Computer Science
and Dept. of  \n  Mathematics and Statistics, James Cook University of
North Queensland. \n  (Also submitted to Technometrics). \n\n  The
data was used with many others for comparing various \n  classifiers.
The classes are separable, though only RDA \n  has achieved 100%
correct classification. \n  (RDA : 100%, QDA 99.4%, LDA 98.9%, 1NN
96.1% (z-transformed data)) \n  (All results using the leave-one-out
technique) \n\n  (2) S. Aeberhard, D. Coomans and O. de Vel, \n  "THE
CLASSIFICATION PERFORMANCE OF RDA" \n  Tech. Rep. no. 92-01, (1992),
Dept. of Computer Science and Dept. of \n  Mathematics and Statistics,
James Cook University of North Queensland. \n  (Also submitted to
Journal of Chemometrics).\n'
```

```
# Вывод зависимых переменных.
data.target
```

```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1,
1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2,
2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
2,
       2, 2])
```

```
# Вывод целевых имен.
data.target_names
```

```
array(['class_0', 'class_1', 'class_2'], dtype='<U7')
```

Конвертация датасета

```
# Теперь конвертируем загруженный набор данных из sklearn в формат
pandas dataframe.

df = pd.DataFrame(data.data, columns=data.feature_names)
type(df)
```

```
pandas.core.frame.DataFrame
```

```
# Выведем конвертированный датасет в виде таблицы.
df.head()
```

```
   alcohol  malic_acid   ash  alcalinity_of_ash  magnesium
total_phenols  \
0    14.23        1.71  2.43               15.6      127.0
2.80
1    13.20        1.78  2.14               11.2      100.0
2.65
2    13.16        2.36  2.67               18.6      101.0
2.80
3    14.37        1.95  2.50               16.8      113.0
3.85
4    13.24        2.59  2.87               21.0      118.0
2.80

   flavanoids  nonflavanoid_phenols  proanthocyanins  color_intensity
hue  \
0        3.06                  0.28             2.29             5.64
1.04
1        2.76                  0.26             1.28             4.38
1.05
2        3.24                  0.30             2.81             5.68
1.03
3        3.49                  0.24             2.18             7.80
0.86
4        2.69                  0.39             1.82             4.32
1.04

   od280/od315_of_diluted_wines  proline
0                          3.92   1065.0
1                          3.40   1050.0
2                          3.17   1185.0
3                          3.45   1480.0
4                          2.93    735.0
```

```python
# Выведем размер датасета - по итогу получилось:
total_count = df.shape[0]
print('Всего строк: {}'.format(total_count))
total_count = df.shape[1]
print('Всего колонок: {}'.format(total_count))
```

```
Всего строк: 178
Всего колонок: 13
```

```python
# Выведем список колонок.
df.dtypes
```

```
alcohol                         float64
malic_acid                      float64
ash                             float64
alcalinity_of_ash               float64
magnesium                       float64
total_phenols                   float64
flavanoids                      float64
nonflavanoid_phenols            float64
proanthocyanins                 float64
color_intensity                 float64
hue                             float64
od280/od315_of_diluted_wines    float64
proline                         float64
dtype: object
```

```python
# Проверил количество пустых значений по колонкам.
for col_empty in df.columns:
    empty_count = df[df[col_empty].isnull()].shape[0]
    print('{} - {}'.format(col_empty, empty_count))
```

```
alcohol - 0
malic_acid - 0
ash - 0
alcalinity_of_ash - 0
magnesium - 0
total_phenols - 0
flavanoids - 0
nonflavanoid_phenols - 0
proanthocyanins - 0
color_intensity - 0
hue - 0
od280/od315_of_diluted_wines - 0
proline - 0
```

Количество пустых значений означает, что все значения по этим колонкам заполнены.

Информация о корреляции признаков

```python
df.corr()
```

```
                               alcohol  malic_acid       ash  \
alcohol                       1.000000    0.094397  0.211545   
malic_acid                    0.094397    1.000000  0.164045   
ash                           0.211545    0.164045  1.000000   
alcalinity_of_ash            -0.310235    0.288500  0.443367   
magnesium                     0.270798   -0.054575  0.286587   
total_phenols                 0.289101   -0.335167  0.128980   
flavanoids                    0.236815   -0.411007  0.115077   
nonflavanoid_phenols         -0.155929    0.292977  0.186230   
proanthocyanins               0.136698   -0.220746  0.009652   
color_intensity               0.546364    0.248985  0.258887   
hue                          -0.071747   -0.561296 -0.074667   
od280/od315_of_diluted_wines  0.072343   -0.368710  0.003911   
proline                       0.643720   -0.192011  0.223626   

                              alcalinity_of_ash  magnesium  
total_phenols  \
alcohol                               -0.310235   0.270798   
0.289101   
malic_acid                             0.288500  -0.054575      -
0.335167   
ash                                    0.443367   0.286587   
0.128980   
alcalinity_of_ash                      1.000000  -0.083333      -
0.321113   
magnesium                             -0.083333   1.000000   
0.214401   
total_phenols                         -0.321113   0.214401   
1.000000   
flavanoids                            -0.351370   0.195784   
0.864564   
nonflavanoid_phenols                   0.361922  -0.256294      -
0.449935   
proanthocyanins                       -0.197327   0.236441   
0.612413   
color_intensity                        0.018732   0.199950      -
0.055136   
hue                                   -0.273955   0.055398   
0.433681   
od280/od315_of_diluted_wines          -0.276769   0.066004   
0.699949   
proline                               -0.440597   0.393351   
0.498115   

                              flavanoids  nonflavanoid_phenols  \
alcohol                         0.236815             -0.155929   
malic_acid                     -0.411007              0.292977   
ash                             0.115077              0.186230   
alcalinity_of_ash              -0.351370              0.361922   
magnesium                       0.195784             -0.256294   
```

```
total_phenols                     0.864564          -0.449935
flavanoids                        1.000000          -0.537900
nonflavanoid_phenols             -0.537900           1.000000
proanthocyanins                   0.652692          -0.365845
color_intensity                  -0.172379           0.139057
hue                               0.543479          -0.262640
od280/od315_of_diluted_wines      0.787194          -0.503270
proline                           0.494193          -0.311385

                              proanthocyanins  color_intensity
hue  \
alcohol                              0.136698         0.546364 -
0.071747
malic_acid                          -0.220746         0.248985 -
0.561296
ash                                  0.009652         0.258887 -
0.074667
alcalinity_of_ash                   -0.197327         0.018732 -
0.273955
magnesium                            0.236441         0.199950
0.055398
total_phenols                        0.612413        -0.055136
0.433681
flavanoids                           0.652692        -0.172379
0.543479
nonflavanoid_phenols                -0.365845         0.139057 -
0.262640
proanthocyanins                      1.000000        -0.025250
0.295544
color_intensity                     -0.025250         1.000000 -
0.521813
hue                                  0.295544        -0.521813
1.000000
od280/od315_of_diluted_wines         0.519067        -0.428815
0.565468
proline                              0.330417         0.316100
0.236183

                              od280/od315_of_diluted_wines   proline
alcohol                                           0.072343  0.643720
malic_acid                                       -0.368710 -0.192011
ash                                               0.003911  0.223626
alcalinity_of_ash                                -0.276769 -0.440597
magnesium                                         0.066004  0.393351
total_phenols                                     0.699949  0.498115
flavanoids                                        0.787194  0.494193
nonflavanoid_phenols                             -0.503270 -0.311385
proanthocyanins                                   0.519067  0.330417
color_intensity                                  -0.428815  0.316100
hue                                               0.565468  0.236183
```

```
od280/od315_of_diluted_wines                      1.000000  0.312761
proline                                           0.312761  1.000000
```

Коэффициент корреляции Пирсона

```
df.corr(method='pearson')
```

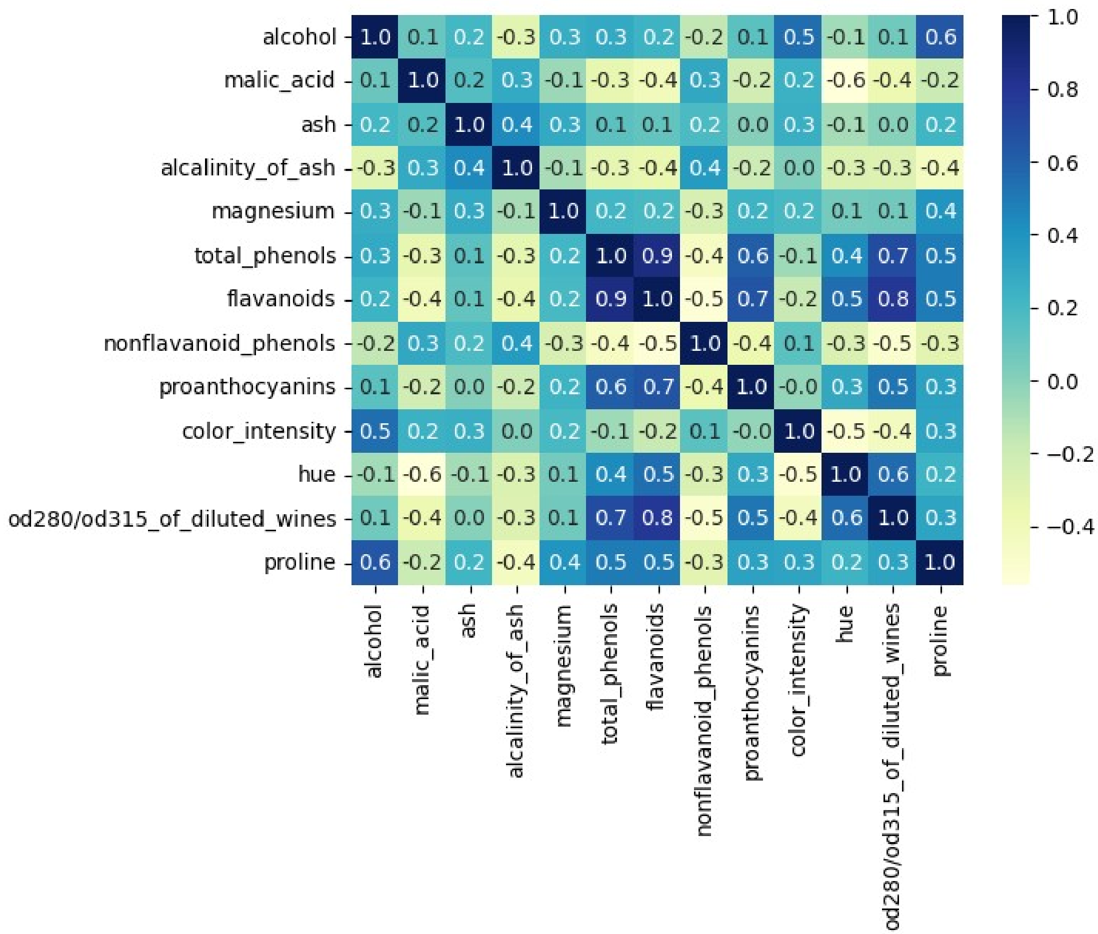
```
                               alcohol  malic_acid       ash  \
alcohol                       1.000000    0.094397  0.211545
malic_acid                    0.094397    1.000000  0.164045
ash                           0.211545    0.164045  1.000000
alcalinity_of_ash            -0.310235    0.288500  0.443367
magnesium                     0.270798   -0.054575  0.286587
total_phenols                 0.289101   -0.335167  0.128980
flavanoids                    0.236815   -0.411007  0.115077
nonflavanoid_phenols         -0.155929    0.292977  0.186230
proanthocyanins               0.136698   -0.220746  0.009652
color_intensity               0.546364    0.248985  0.258887
hue                          -0.071747   -0.561296 -0.074667
od280/od315_of_diluted_wines  0.072343   -0.368710  0.003911
proline                       0.643720   -0.192011  0.223626

                              alcalinity_of_ash  magnesium
total_phenols  \
alcohol                               -0.310235   0.270798
0.289101
malic_acid                             0.288500  -0.054575      -
0.335167
ash                                    0.443367   0.286587
0.128980
alcalinity_of_ash                      1.000000  -0.083333      -
0.321113
magnesium                             -0.083333   1.000000
0.214401
total_phenols                         -0.321113   0.214401
1.000000
flavanoids                            -0.351370   0.195784
0.864564
nonflavanoid_phenols                   0.361922  -0.256294      -
0.449935
proanthocyanins                       -0.197327   0.236441
0.612413
color_intensity                        0.018732   0.199950      -
0.055136
hue                                   -0.273955   0.055398
0.433681
od280/od315_of_diluted_wines          -0.276769   0.066004
0.699949
proline                               -0.440597   0.393351
0.498115
```

```
                              flavanoids  nonflavanoid_phenols  \
alcohol                         0.236815             -0.155929
malic_acid                     -0.411007              0.292977
ash                             0.115077              0.186230
alcalinity_of_ash              -0.351370              0.361922
magnesium                       0.195784             -0.256294
total_phenols                   0.864564             -0.449935
flavanoids                      1.000000             -0.537900
nonflavanoid_phenols           -0.537900              1.000000
proanthocyanins                 0.652692             -0.365845
color_intensity                -0.172379              0.139057
hue                             0.543479             -0.262640
od280/od315_of_diluted_wines    0.787194             -0.503270
proline                         0.494193             -0.311385

                              proanthocyanins  color_intensity
hue  \
alcohol                              0.136698         0.546364 -
0.071747
malic_acid                          -0.220746         0.248985 -
0.561296
ash                                  0.009652         0.258887 -
0.074667
alcalinity_of_ash                   -0.197327         0.018732 -
0.273955
magnesium                            0.236441         0.199950
0.055398
total_phenols                        0.612413        -0.055136
0.433681
flavanoids                           0.652692        -0.172379
0.543479
nonflavanoid_phenols                -0.365845         0.139057 -
0.262640
proanthocyanins                      1.000000        -0.025250
0.295544
color_intensity                     -0.025250         1.000000 -
0.521813
hue                                  0.295544        -0.521813
1.000000
od280/od315_of_diluted_wines         0.519067        -0.428815
0.565468
proline                              0.330417         0.316100
0.236183

                              od280/od315_of_diluted_wines   proline
alcohol                                           0.072343  0.643720
malic_acid                                       -0.368710 -0.192011
ash                                               0.003911  0.223626
alcalinity_of_ash                                -0.276769 -0.440597
magnesium                                         0.066004  0.393351
```

```
total_phenols                                         0.699949  0.498115
flavanoids                                            0.787194  0.494193
nonflavanoid_phenols                                 -0.503270 -0.311385
proanthocyanins                                       0.519067  0.330417
color_intensity                                      -0.428815  0.316100
hue                                                   0.565468  0.236183
od280/od315_of_diluted_wines                          1.000000  0.312761
proline                                               0.312761  1.000000
```

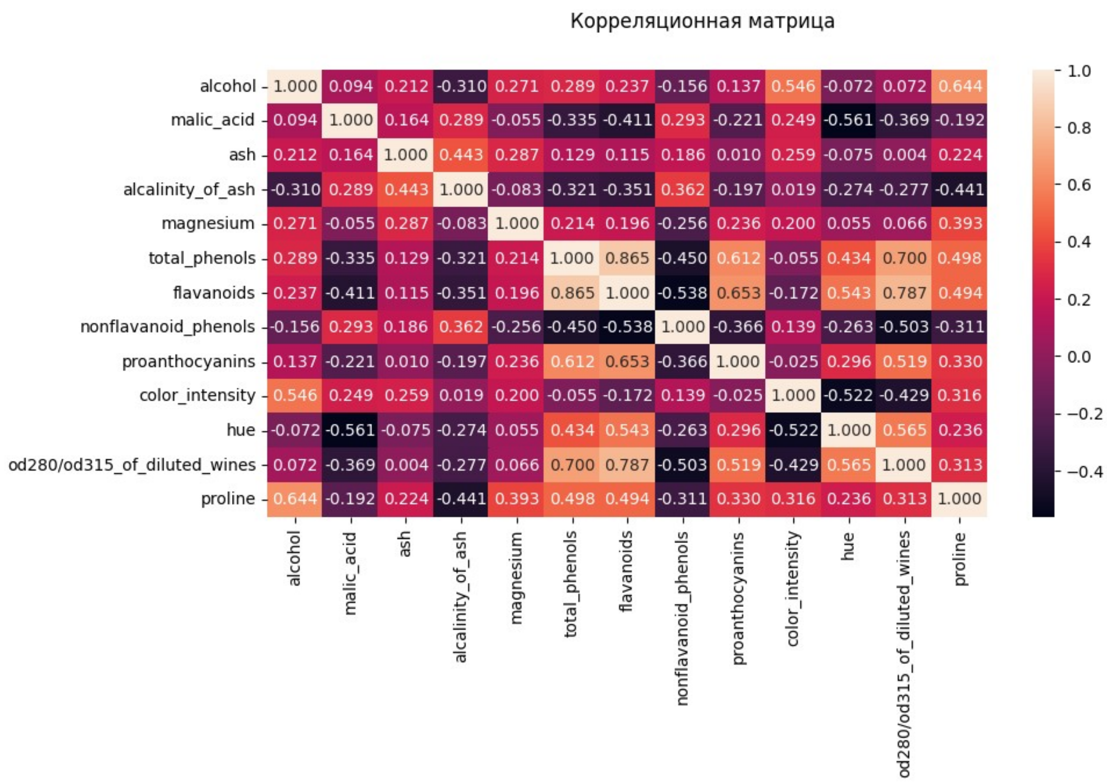
```python
sns.heatmap(df.corr(), cmap='YlGnBu', annot=True, fmt='.1f')
```

```
<AxesSubplot:>
```




```python
fig, ax = plt.subplots(1, 1, sharex='col', sharey='row',
figsize=(10,5))
fig.suptitle('Корреляционная матрица')
sns.heatmap(df.corr(), ax=ax, annot=True, fmt='.3f')
```

```
<AxesSubplot:>
```



Визуальное исследование датасета

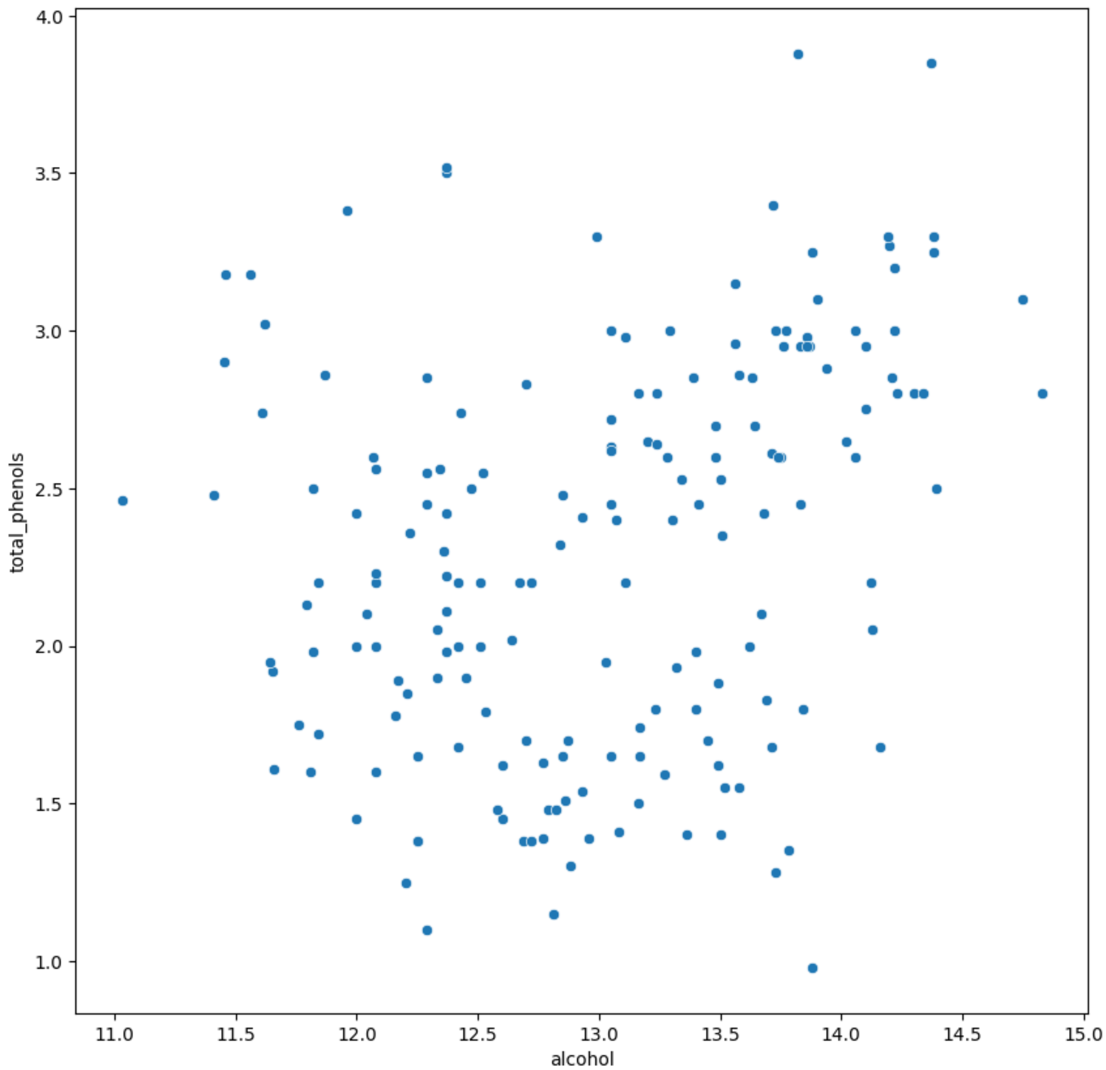
Диаграмма рассеяния

```
# Вывел основные статистические характеристки набора данных этого датасета.
df.describe()

          alcohol  malic_acid         ash  alcalinity_of_ash
magnesium  \
count  178.000000  178.000000  178.000000         178.000000
178.000000
mean    13.000618    2.336348    2.366517          19.494944
99.741573
std      0.811827    1.117146    0.274344           3.339564
14.282484
min     11.030000    0.740000    1.360000          10.600000
70.000000
25%     12.362500    1.602500    2.210000          17.200000
88.000000
50%     13.050000    1.865000    2.360000          19.500000
98.000000
75%     13.677500    3.082500    2.557500          21.500000
107.000000
max     14.830000    5.800000    3.230000          30.000000
162.000000
```

```
       total_phenols  flavanoids  nonflavanoid_phenols  
proanthocyanins  \
count     178.000000  178.000000            178.000000  
178.000000
mean        2.295112    2.029270              0.361854  
1.590899
std         0.625851    0.998859              0.124453  
0.572359
min         0.980000    0.340000              0.130000  
0.410000
25%         1.742500    1.205000              0.270000  
1.250000
50%         2.355000    2.135000              0.340000  
1.555000
75%         2.800000    2.875000              0.437500  
1.950000
max         3.880000    5.080000              0.660000  
3.580000

       color_intensity         hue  od280/od315_of_diluted_wines  
proline
count       178.000000  178.000000                    178.000000  
178.000000
mean          5.058090    0.957449                      2.611685  
746.893258
std           2.318286    0.228572                      0.709990  
314.907474
min           1.280000    0.480000                      1.270000  
278.000000
25%           3.220000    0.782500                      1.937500  
500.500000
50%           4.690000    0.965000                      2.780000  
673.500000
75%           6.200000    1.120000                      3.170000  
985.000000
max          13.000000    1.710000                      4.000000  
1680.000000
```

```
fig, ax = plt.subplots(figsize=(10,10))
sns.scatterplot(ax=ax, x='alcohol', y='total_phenols', data=df)
```

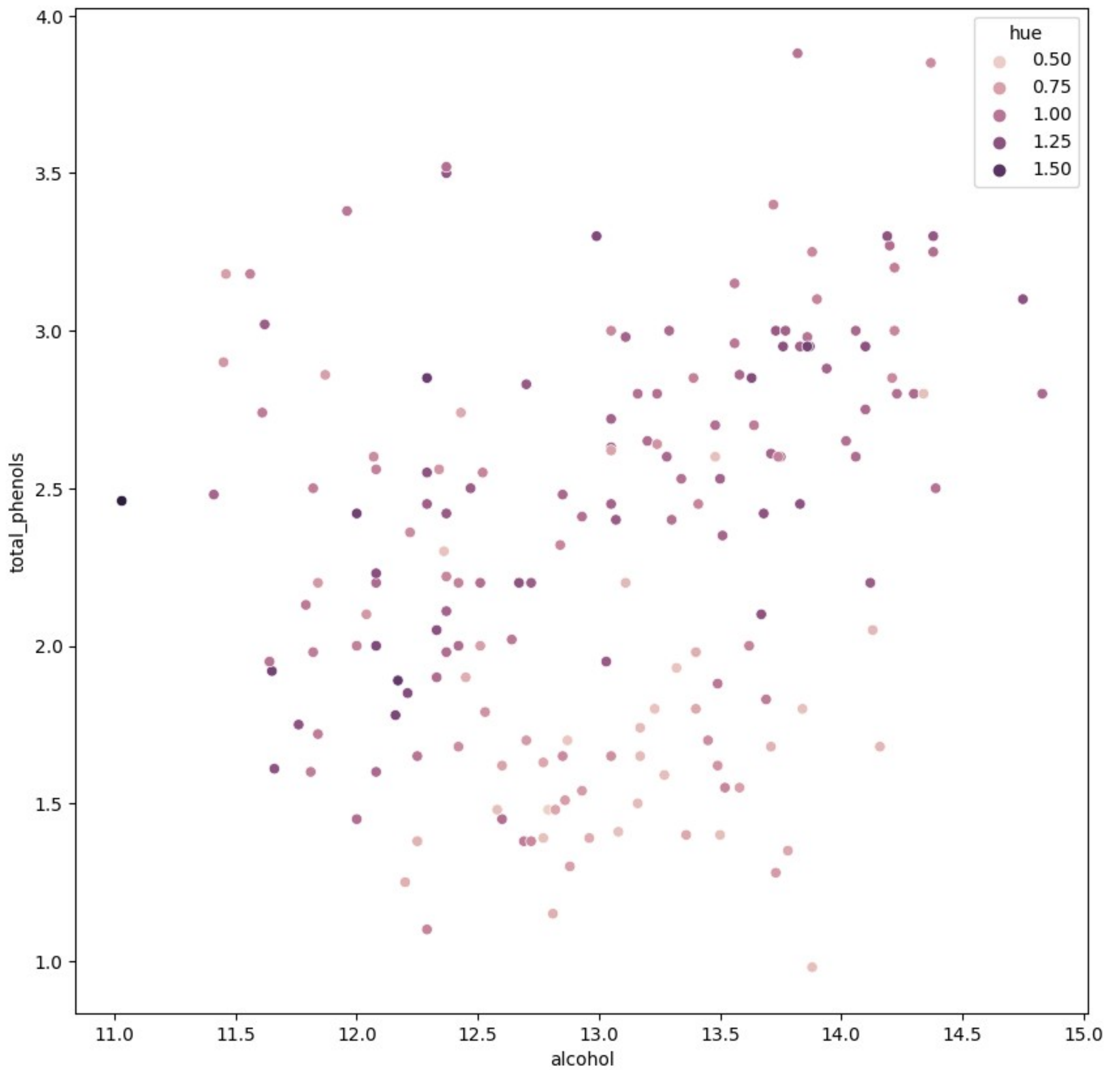
```
<AxesSubplot:xlabel='alcohol', ylabel='total_phenols'>
```



Данная диаграмма показывает количество фенолов в каждом проценте вина.

```
fig, ax = plt.subplots(figsize=(10,10))
sns.scatterplot(ax=ax, x='alcohol', y='total_phenols', data=df,
hue='hue')
```

```
<AxesSubplot:xlabel='alcohol', ylabel='total_phenols'>
```

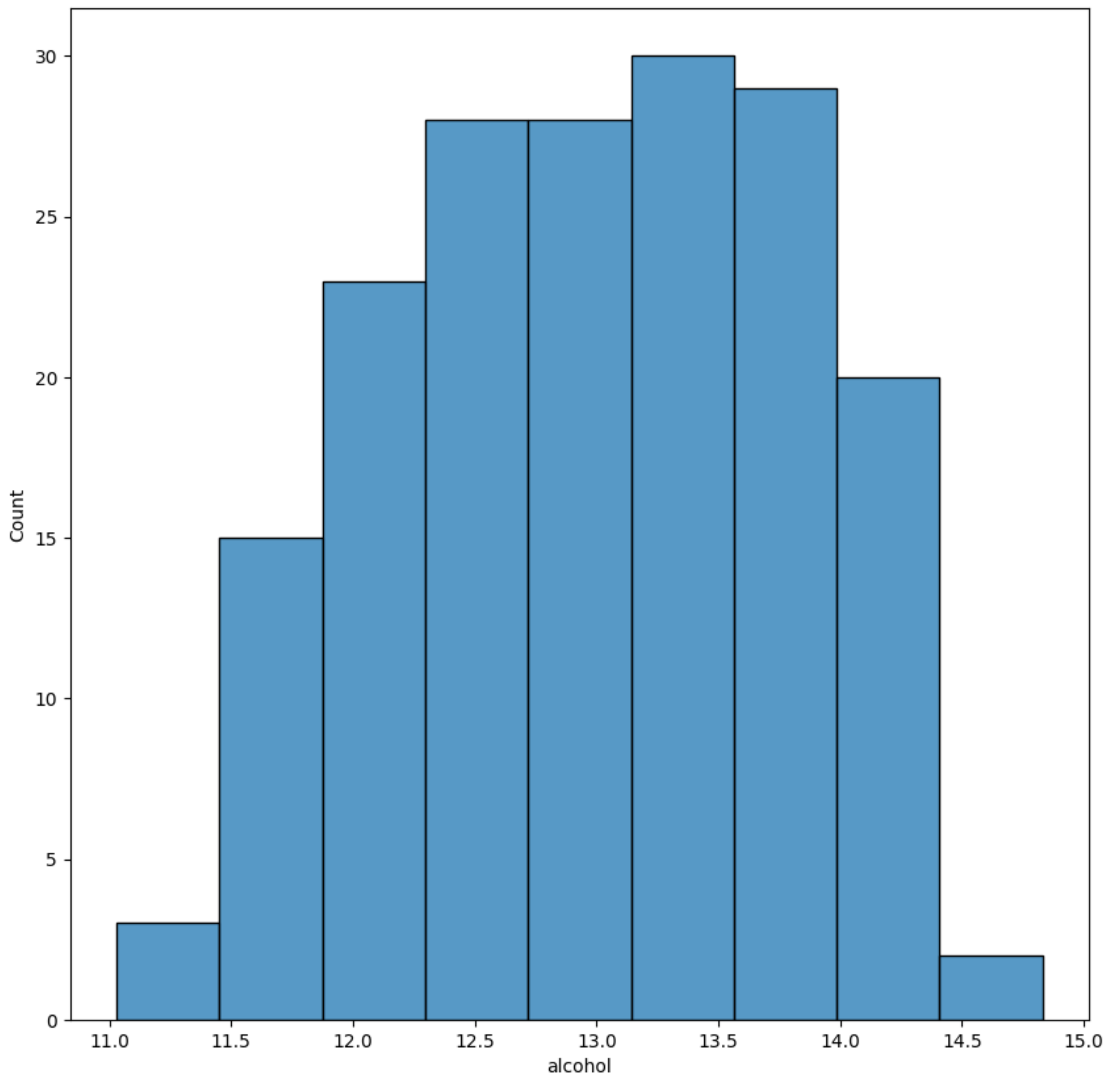


Такая же диаграмма показывает количество фенолов в каждом проценте вина, но еще добавили "hue", т.е. в каждой точке можем рассмотреть оттеннок конкретного вина.

Гистограмма

```
fig, ax = plt.subplots(figsize=(10,10))
sns.histplot(df['alcohol'])
```

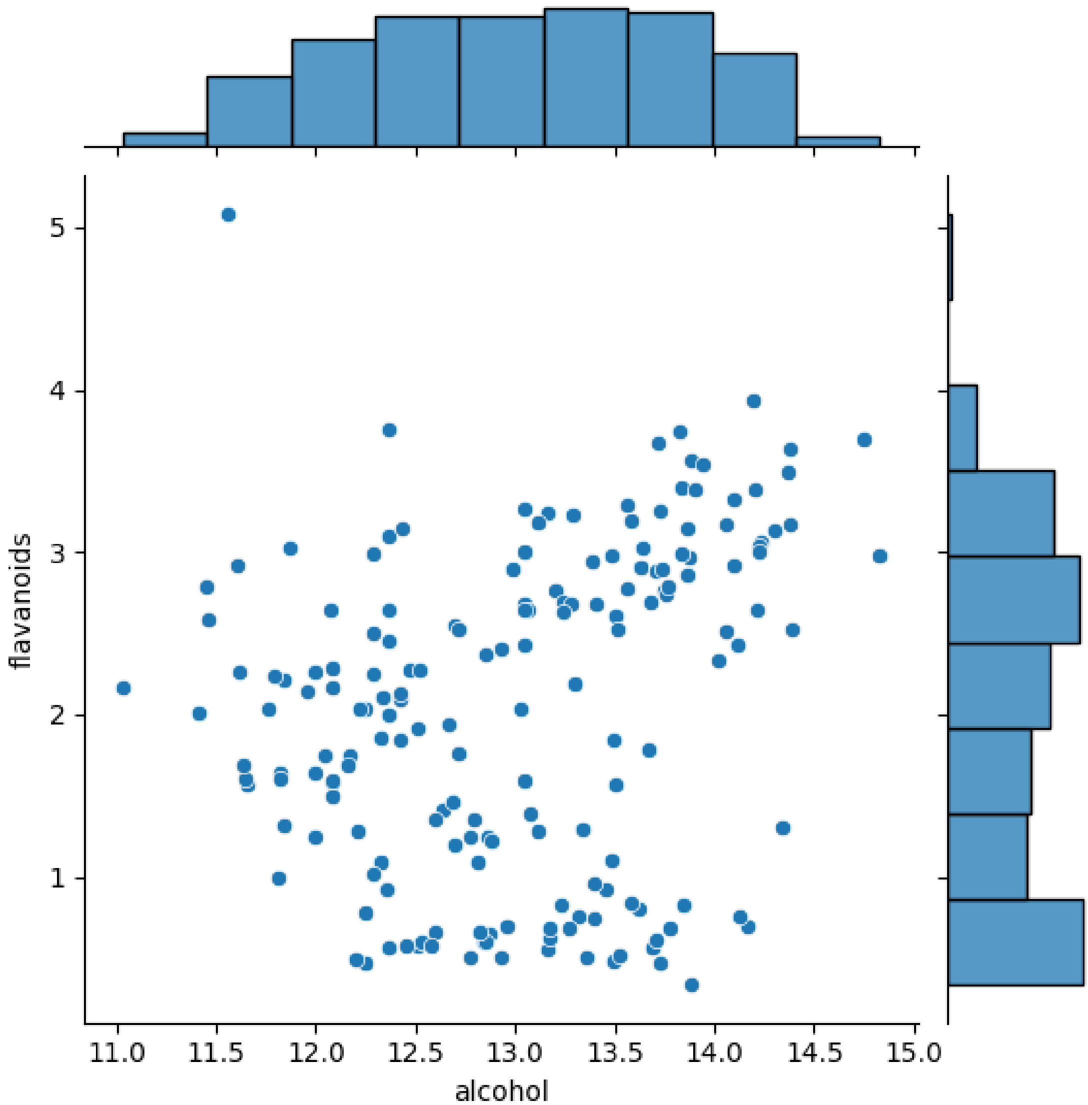
```
<AxesSubplot:xlabel='alcohol', ylabel='Count'>
```



Jointplot

```
sns.jointplot(x='alcohol', y='flavanoids', data=df)
```

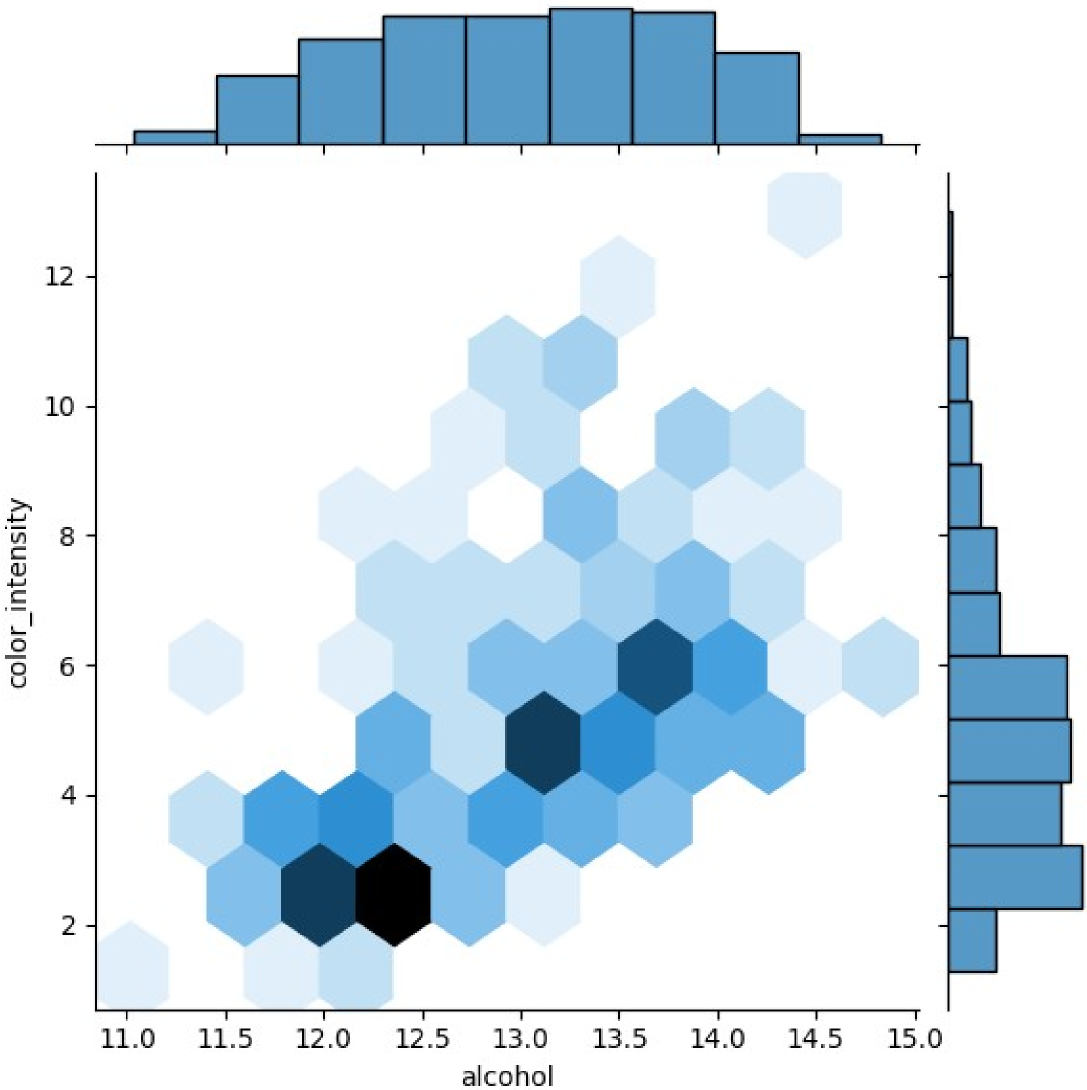
```
<seaborn.axisgrid.JointGrid at 0x1a8e080fbb0>
```



Комбинация гистограмм и диаграмм рассеивания.

```
sns.jointplot(x='alcohol', y='color_intensity', data=df, kind="hex")
```

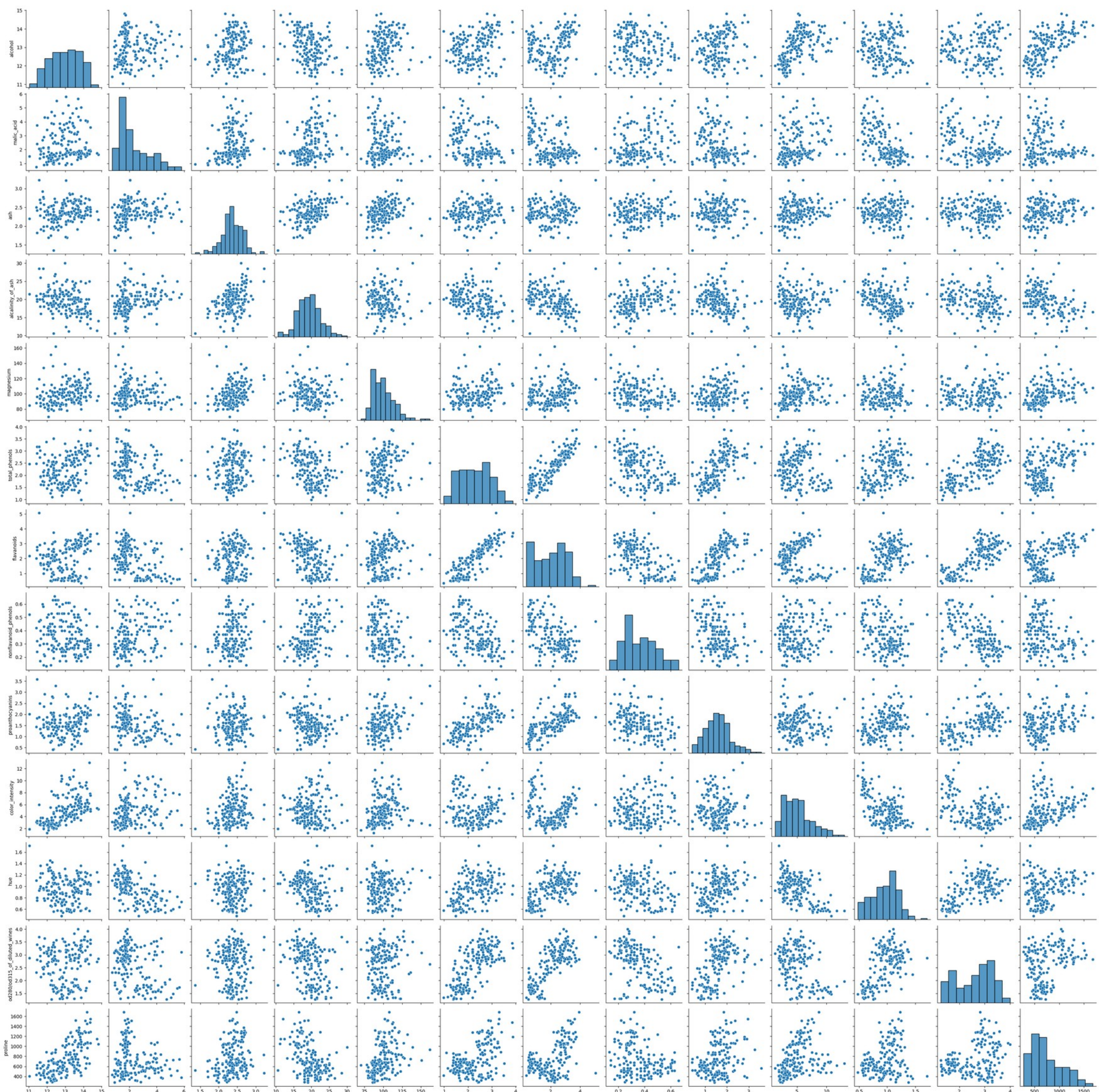
```
<seaborn.axisgrid.JointGrid at 0x1a8e0e135e0>
```



"Парные диаграммы"

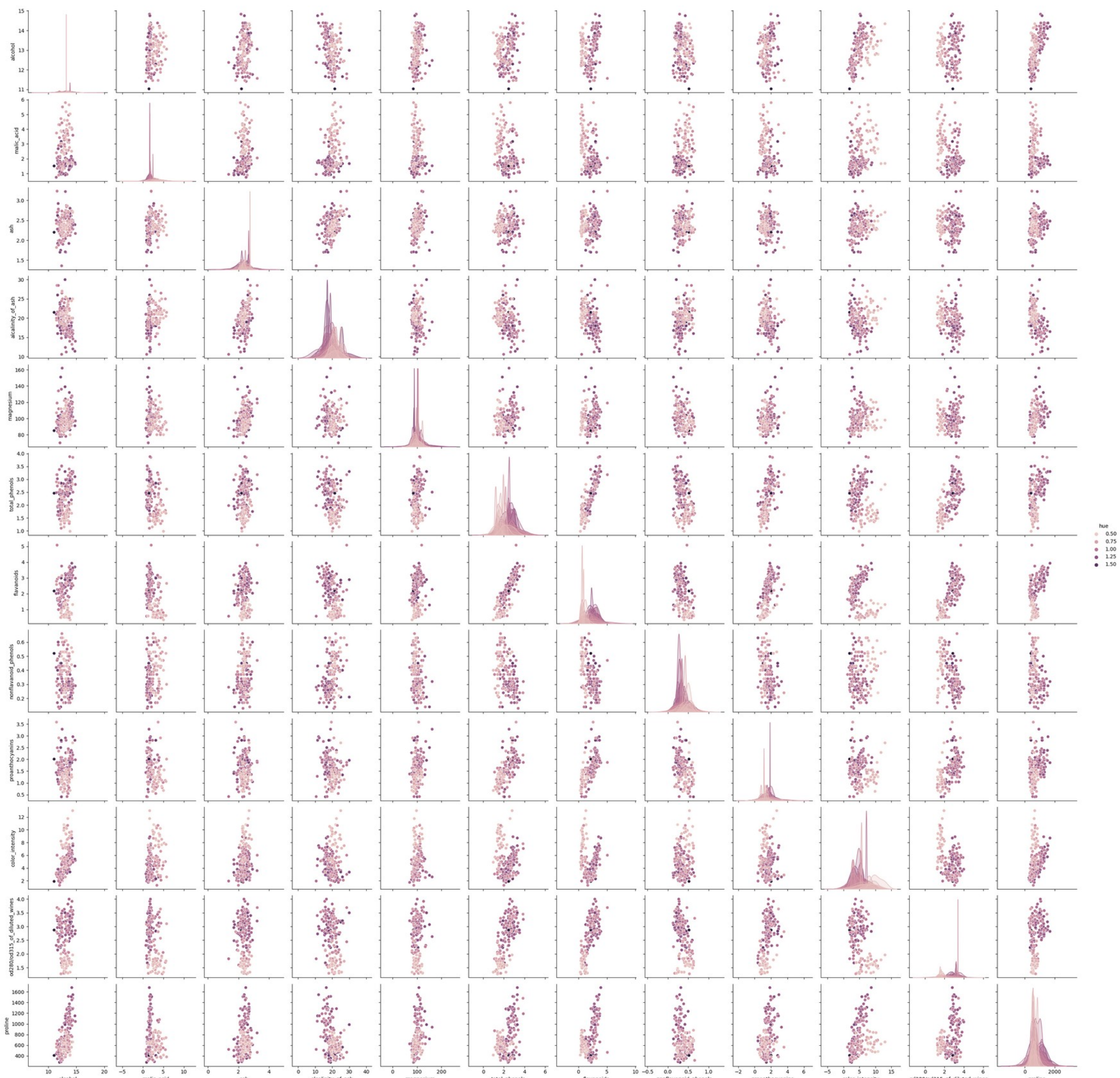
```
sns.pairplot(df)
```

```
<seaborn.axisgrid.PairGrid at 0x1a8ea7cebf0>
```



```
sns.pairplot(df, hue="hue")
```

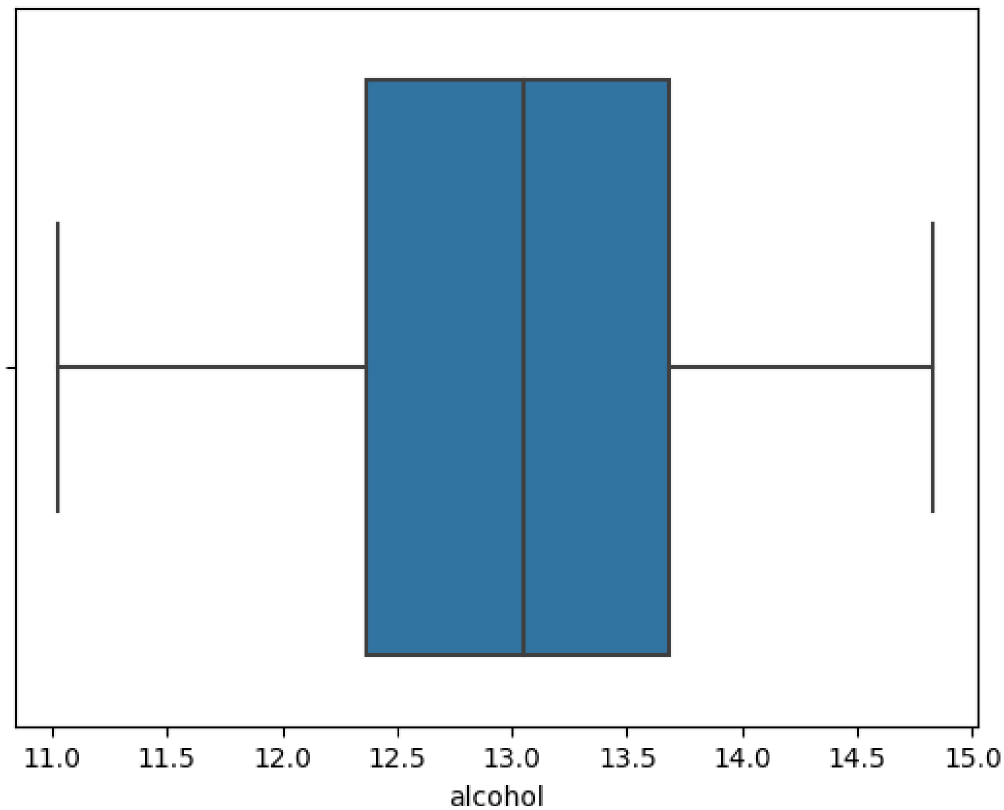
<seaborn.axisgrid.PairGrid at 0x1a8f314e650>



"Ящик с усами"

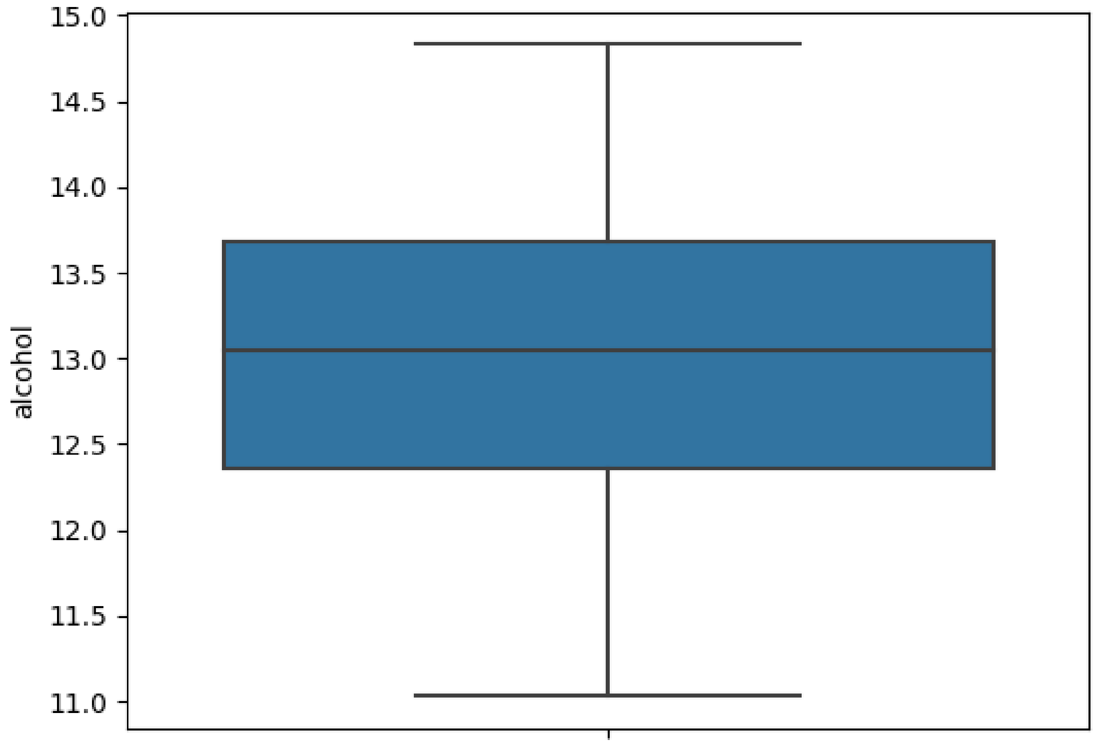
```
# по оси абсцисс.
sns.boxplot(x=df['alcohol'])
```

```
<AxesSubplot:xlabel='alcohol'>
```




```
# По оси ординат
sns.boxplot(y=df['alcohol'])
```

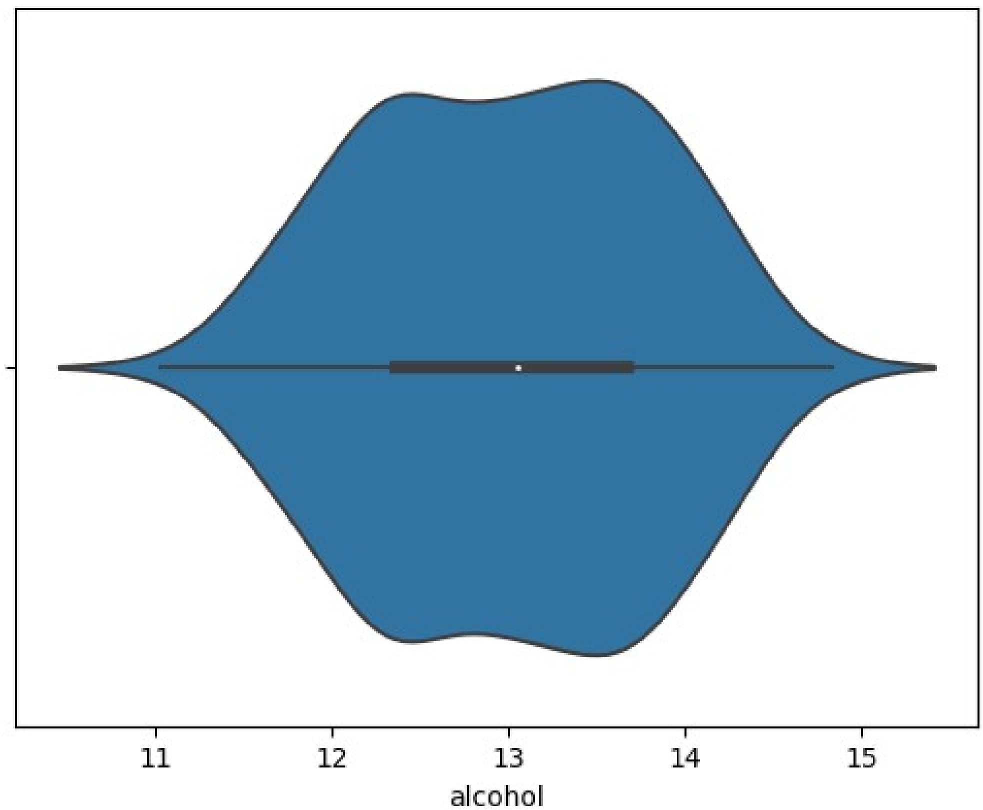
```
<AxesSubplot:ylabel='alcohol'>
```



Скрипичная диаграмма

```
sns.violinplot(x=df['alcohol'])
```

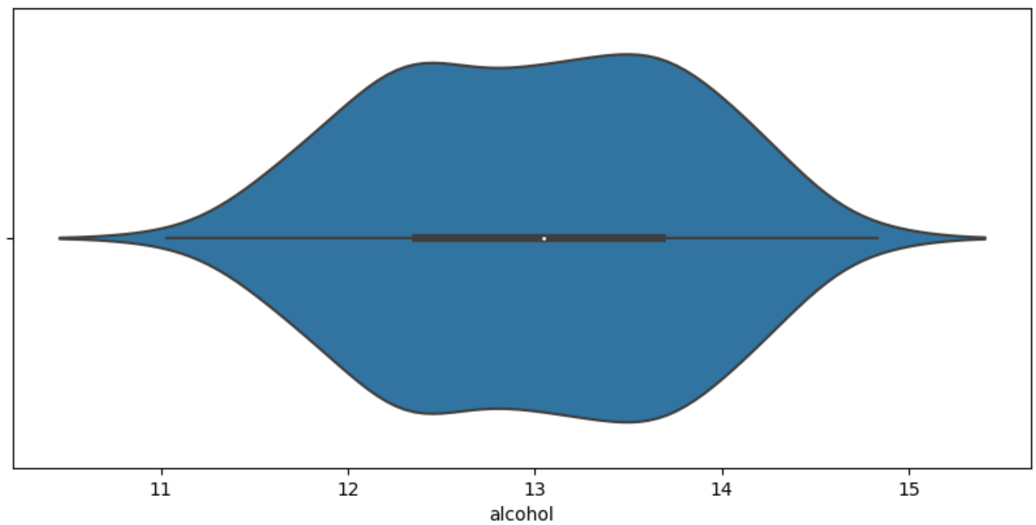
```
<AxesSubplot:xlabel='alcohol'>
```

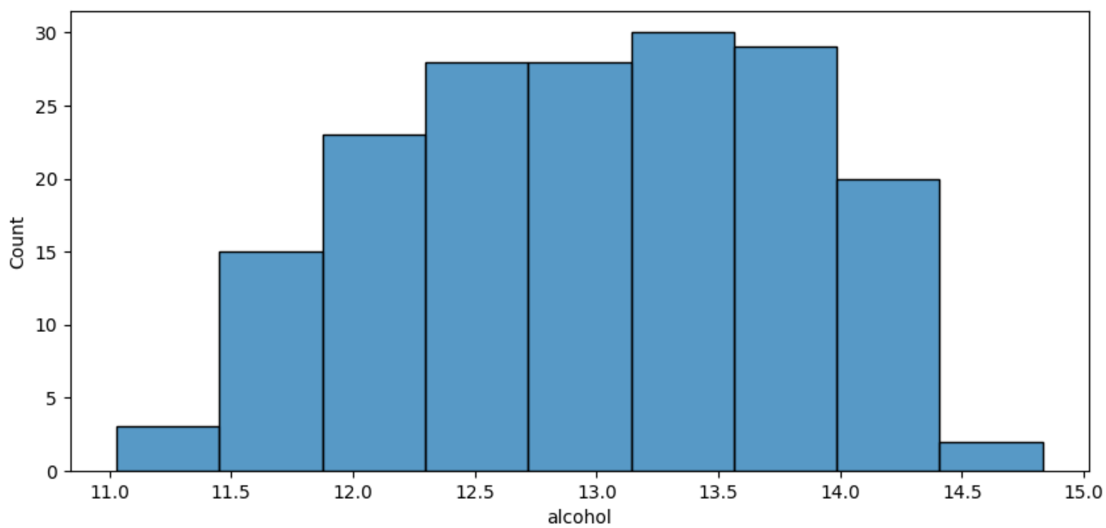


Скрипичная диаграмма показывает распределение плотности по краям диаграммы.

```
fig, ax = plt.subplots(2, 1, figsize=(10,10))
sns.violinplot(ax=ax[0], x=df['alcohol'])
sns.histplot(df['alcohol'])
```

```
<AxesSubplot:xlabel='alcohol', ylabel='Count'>
```





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