Министерство образования и науки 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)

Методы машинного обучения

Отчёт по лабораторной работе № 1

«Разведочный анализ данных. Исследование и визуализация данных.»

Выполнил:
студент группы ИУ5 – 23М Крутов Т.Ю.

Преподаватель: Гапанюк Ю.Е.

2020г.

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

```
dataset = pd.read_csv('wine.csv', sep = ",")
```

```
dataset.head()
```

	Wine	Alcohol	Malic.acid	Ash	Acl	Mg	Phenols	Flavanoids	Nonflavanoid.phenc
0	1	14.23	1.71	2.43	15.6	127	2.80	3.06	0.
1	1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.
2	1	13.16	2.36	2.67	18.6	101	2.80	3.24	0.
3	1	14.37	1.95	2.50	16.8	113	3.85	3.49	0.
4	1	13.24	2.59	2.87	21.0	118	2.80	2.69	0.

```
dataset.shape
```

```
(178, 14)
```

```
dataset.dtypes
```

```
Wine                      int64
Alcohol                 float64
Malic.acid              float64
Ash                     float64
Acl                     float64
Mg                        int64
Phenols                 float64
Flavanoids              float64
Nonflavanoid.phenols    float64
Proanth                 float64
Color.int               float64
Hue                     float64
OD                      float64
Proline                   int64
dtype: object
```

```
for col in dataset.columns:
  temp_null_count = dataset[dataset[col].isnull()].shape[0]
  print('{} - {}'.format(col, temp_null_count))
```

```
Wine - 0
Alcohol - 0
Malic.acid - 0
Ash - 0
Acl - 0
Mg - 0
Phenols - 0
Flavanoids - 0
Nonflavanoid.phenols - 0
Proanth - 0
Color.int - 0
Hue - 0
OD - 0
Proline - 0
```

```
dataset.describe()
```

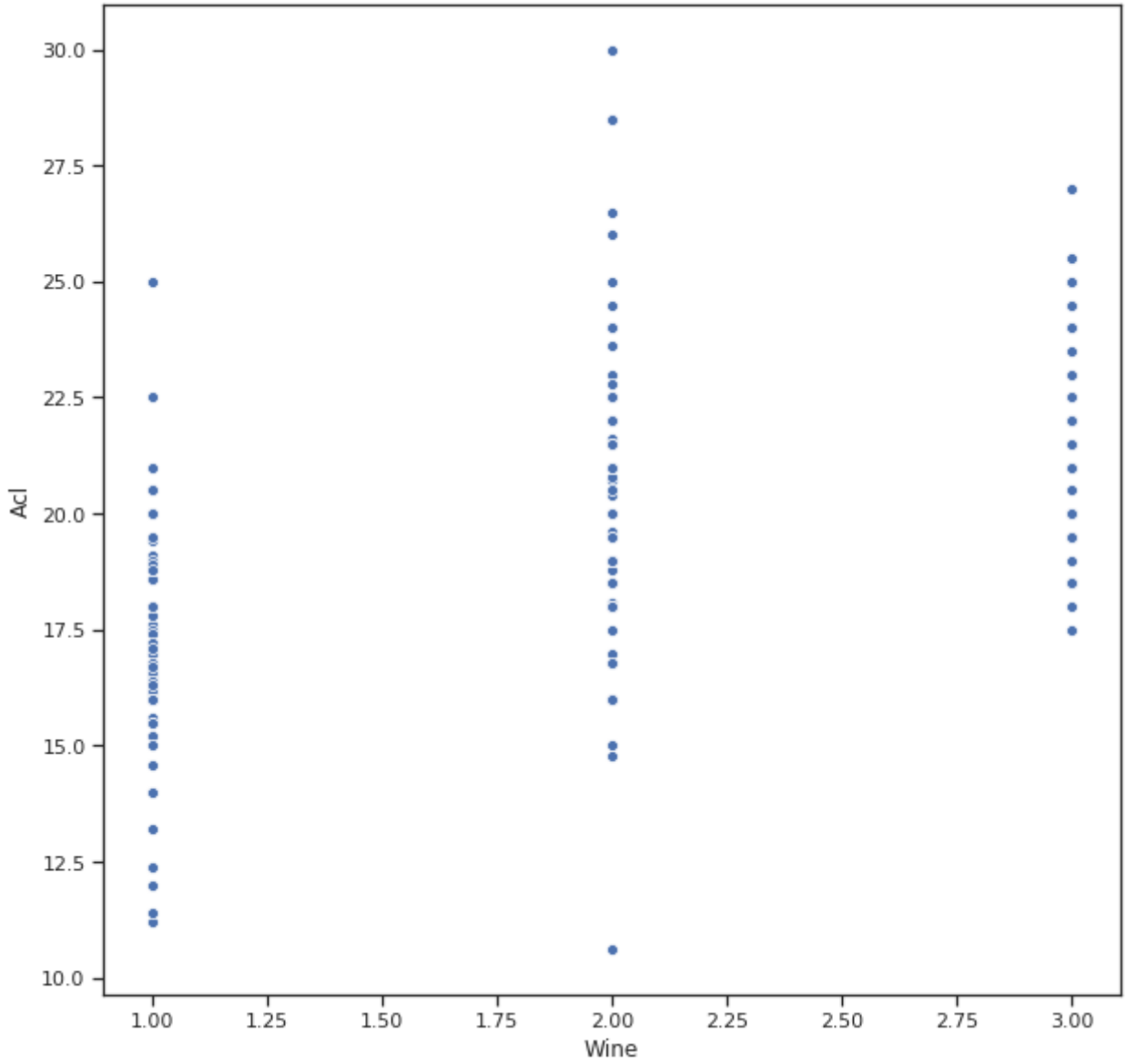
	Wine	Alcohol	Malic.acid	Ash	Acl	Mg	Pheno
count	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.0000
mean	1.938202	13.000618	2.336348	2.366517	19.494944	99.741573	2.2951
std	0.775035	0.811827	1.117146	0.274344	3.339564	14.282484	0.6258
min	1.000000	11.030000	0.740000	1.360000	10.600000	70.000000	0.9800
25%	1.000000	12.362500	1.602500	2.210000	17.200000	88.000000	1.7425
50%	2.000000	13.050000	1.865000	2.360000	19.500000	98.000000	2.3550
75%	3.000000	13.677500	3.082500	2.557500	21.500000	107.000000	2.8000
max	3.000000	14.830000	5.800000	3.230000	30.000000	162.000000	3.8800

```
dataset['Acl'].unique()
```

```
array([15.6, 11.2, 18.6, 16.8, 21. , 15.2, 14.6, 17.6, 14. , 16. , 18. ,
       11.4, 12. , 17.2, 20. , 16.5, 16.6, 17.8, 25. , 16.1, 17. , 19.4,
       22.5, 19.1, 19.5, 19. , 20.5, 15.5, 13.2, 16.2, 18.8, 15. , 17.5,
       18.9, 17.4, 12.4, 17.1, 16.4, 16.3, 16.7, 10.6, 18.1, 19.6, 20.4,
       24. , 30. , 14.8, 23. , 22.8, 26. , 21.6, 23.6, 18.5, 22. , 20.7,
       21.5, 20.8, 28.5, 26.5, 24.5, 23.5, 25.5, 27. ])
```

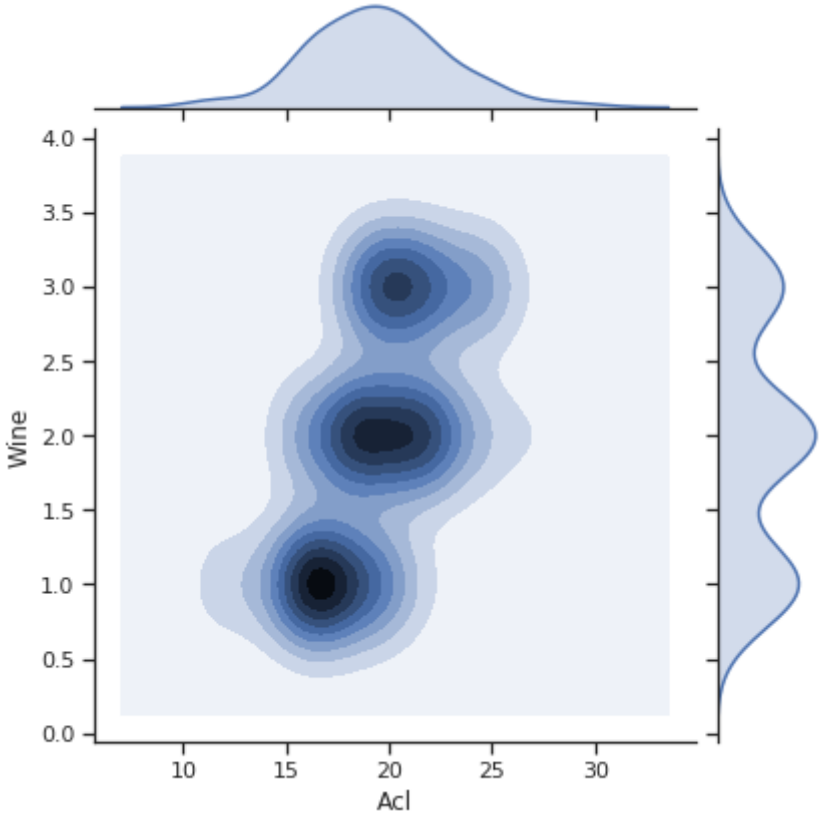
```
fig, ax = plt.subplots(figsize=(10,10))
sns.scatterplot(ax=ax, x='Wine', y='Acl', data=dataset)
#не работает
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f2dde51b2b0>
```




```
sns.jointplot(x = 'Acl', y = 'Wine', data= dataset, kind = 'kde')
```

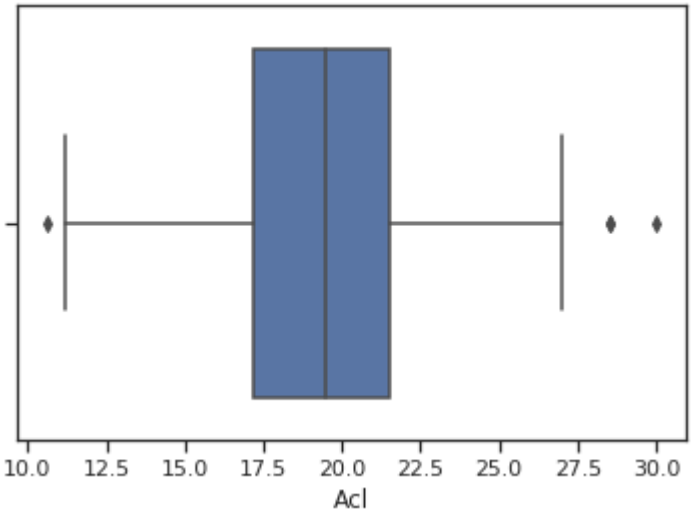
```
<seaborn.axisgrid.JointGrid at 0x7f60222ec320>
```




```
sns.jointplot(x = )
```

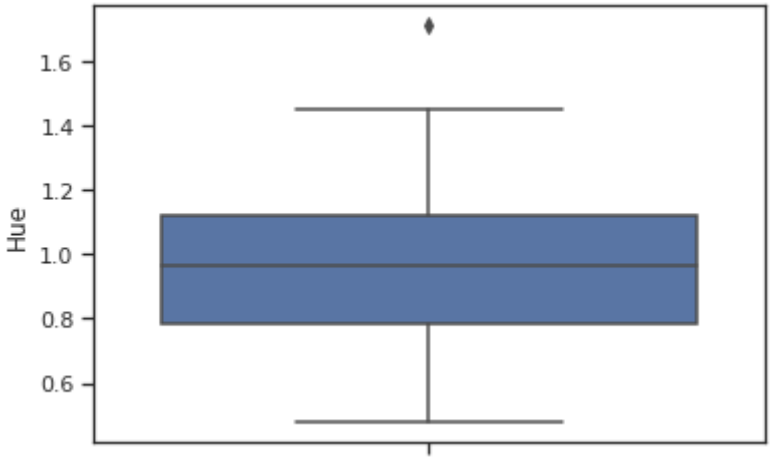
```
sns.boxplot(x=dataset['Acl'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f601ee44c50>
```



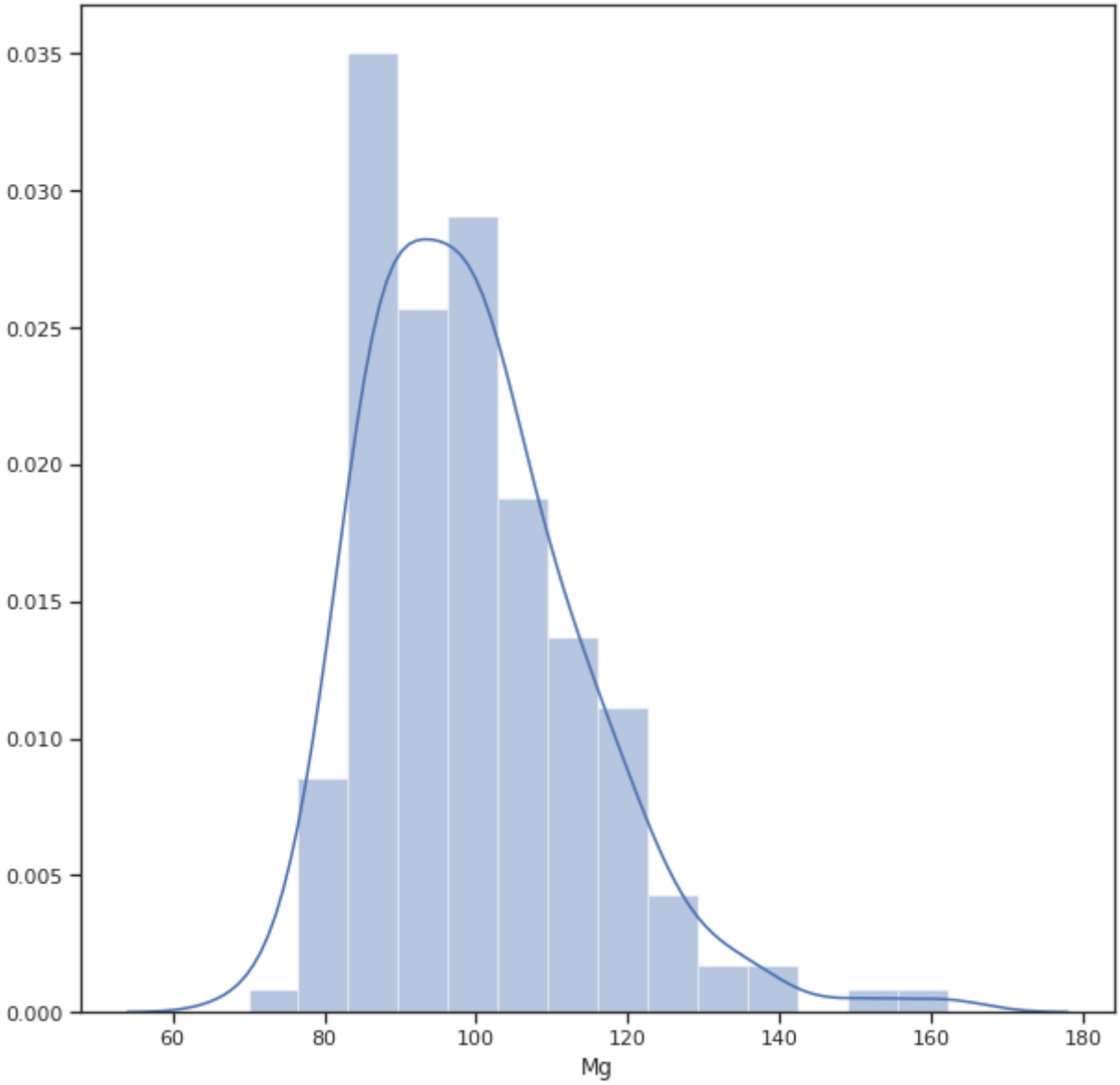

```
sns.boxplot(y=dataset['Hue'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f601db71ef0>
```



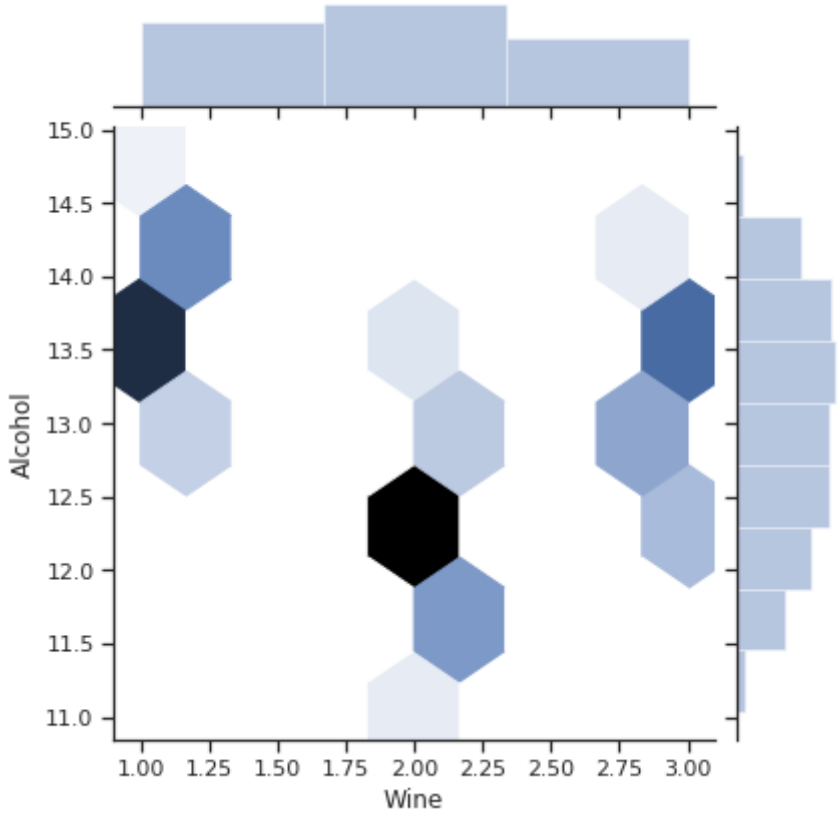

```
fig, ax = plt.subplots(figsize=(10,10))
sns.distplot(dataset['Mg'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f2dda716320>
```




```
sns.jointplot(x='Wine', y='Alcohol', data=dataset, kind="hex")
```

```
<seaborn.axisgrid.JointGrid at 0x7f2dda749748>
```

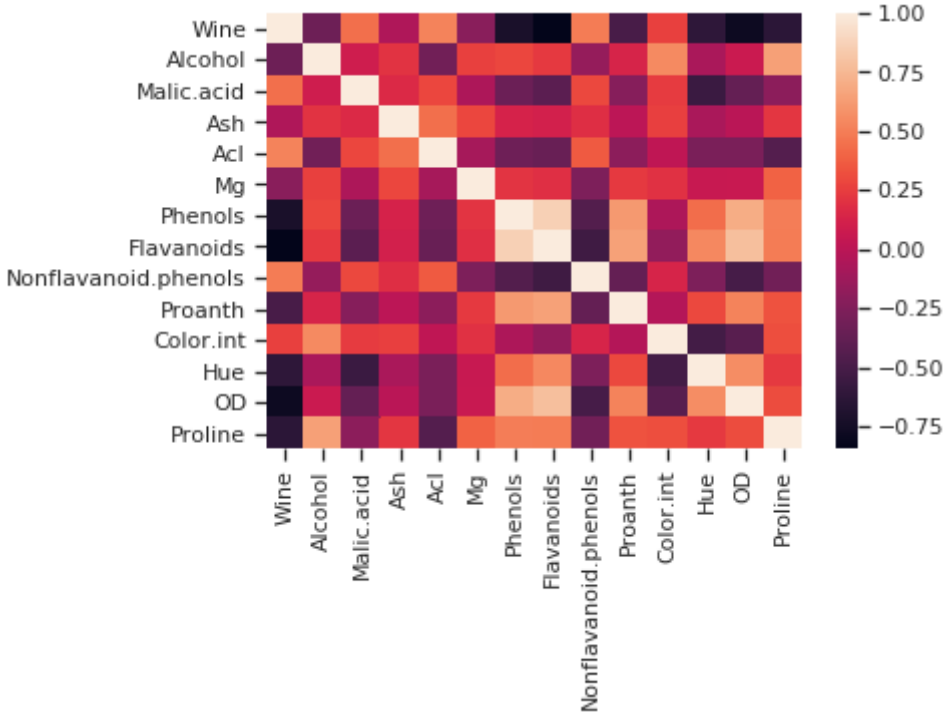



```
dataset.corr()
```

	Wine	Alcohol	Malic.acid	Ash	Acl	Mg
Wine	1.000000	-0.328222	0.437776	-0.049643	0.517859	-0.209179
Alcohol	-0.328222	1.000000	0.094397	0.211545	-0.310235	0.270798
Malic.acid	0.437776	0.094397	1.000000	0.164045	0.288500	-0.054575
Ash	-0.049643	0.211545	0.164045	1.000000	0.443367	0.286587
Acl	0.517859	-0.310235	0.288500	0.443367	1.000000	-0.083333
Mg	-0.209179	0.270798	-0.054575	0.286587	-0.083333	1.000000
Phenols	-0.719163	0.289101	-0.335167	0.128980	-0.321113	0.214401
Flavanoids	-0.847498	0.236815	-0.411007	0.115077	-0.351370	0.195784
Nonflavanoid.phenols	0.489109	-0.155929	0.292977	0.186230	0.361922	-0.256294
Proanth	-0.499130	0.136698	-0.220746	0.009652	-0.197327	0.236441
Color.int	0.265668	0.546364	0.248985	0.258887	0.018732	0.199950
Hue	-0.617369	-0.071747	-0.561296	-0.074667	-0.273955	0.055398
OD	-0.788230	0.072343	-0.368710	0.003911	-0.276769	0.066004
Proline	-0.633717	0.643720	-0.192011	0.223626	-0.440597	0.393351

```
sns.heatmap(dataset.corr())
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f601e758b38>
```




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

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f2dd8e5b978>
```