In [52]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.preprocessing import LabelEncoder
```

In [53]:

```
df = pd.read_csv('D:\\Ботва\\Магистратура\\2сем\\ММО\\РК\\winequality-red.csv')
df
```

Out[53]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol
0	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4
1	7.8	0.880	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8
2	7.8	0.760	0.04	2.3	NaN	15.0	54.0	0.9970	3.26	0.65	9.8
3	11.2	0.280	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8
4	7.4	0.700	0.00	1.9	NaN	11.0	34.0	0.9978	3.51	0.56	9.4
5	7.4	0.660	0.00	1.8	0.075	13.0	40.0	0.9978	3.51	0.56	9.4
6	7.9	0.600	0.06	1.6	0.069	15.0	59.0	0.9964	3.30	0.46	9.4
7	7.3	0.650	0.00	1.2	0.065	15.0	21.0	0.9946	3.39	0.47	10.0
8	7.8	0.580	0.02	2.0	NaN	9.0	18.0	0.9968	3.36	0.57	9.5
9	7.5	0.500	0.36	6.1	0.071	17.0	102.0	0.9978	3.35	0.80	10.5
10	6.7	0.580	0.08	1.8	0.097	15.0	65.0	0.9959	3.28	0.54	9.2
11	7.5	0.500	0.36	6.1	NaN	17.0	102.0	0.9978	3.35	0.80	10.5
12	5.6	0.615	0.00	1.6	0.089	16.0	59.0	0.9943	3.58	0.52	9.9
13	7.8	0.610	0.29	1.6	0.114	9.0	29.0	0.9974	3.26	1.56	9.1
14	8.9	0.620	0.18	3.8	0.176	52.0	145.0	0.9986	3.16	0.88	9.2
15	8.9	0.620	0.19	3.9	0.170	51.0	148.0	0.9986	3.17	0.93	9.2
16	8.5	0.280	0.56	1.8	0.092	35.0	103.0	0.9969	3.30	0.75	10.5
17	8.1	0.560	0.28	1.7	0.368	16.0	56.0	0.9968	3.11	1.28	9.3
18	7.4	0.590	0.08	4.4	NaN	6.0	29.0	0.9974	3.38	0.50	9.0
19	7.9	0.320	0.51	1.8	0.341	17.0	56.0	0.9969	3.04	1.08	9.2
20	8.9	0.220	0.48	1.8	0.077	29.0	60.0	0.9968	3.39	0.53	9.4
21	7.6	0.390	0.31	2.3	0.082	23.0	71.0	0.9982	3.52	0.65	9.7
22	7.9	0.430	0.21	1.6	0.106	10.0	37.0	0.9966	3.17	0.91	9.5
23	8.5	0.490	0.11	2.3	NaN	9.0	67.0	0.9968	3.17	0.53	9.4
24	6.9	0.400	0.14	2.4	0.085	21.0	40.0	0.9968	3.43	0.63	9.7
25	6.3	0.390	0.16	1.4	NaN	11.0	23.0	0.9955	3.34	0.56	9.3
26	7.6	0.410	0.24	1.8	0.080	4.0	11.0	0.9962	3.28	0.59	9.5
27	7.9	0.430	0.21	1.6	0.106	10.0	37.0	0.9966	3.17	0.91	9.5

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol
28	7.1	0.710	0.00	1.9	0.080	14.0	35.0	0.9972	3.47	0.55	9.4
29	7.8	0.645	0.00	2.0	NaN	8.0	16.0	0.9964	3.38	0.59	9.8
30	6.7	0.675	0.07	2.4	0.089	17.0	82.0	0.9958	3.35	0.54	10.1
31	6.9	0.685	0.00	2.5	0.105	22.0	37.0	0.9966	3.46	0.57	10.6
32	8.3	0.655	0.12	2.3	0.083	15.0	113.0	0.9966	3.17	0.66	9.8
33	6.9	0.605	0.12	10.7	0.073	40.0	83.0	0.9993	3.45	0.52	9.4
34	5.2	0.320	0.25	1.8	NaN	13.0	50.0	0.9957	3.38	0.55	9.2
35	7.8	0.645	0.00	5.5	0.086	5.0	18.0	0.9986	3.40	0.55	9.6
36	7.8	0.600	0.14	2.4	0.086	3.0	15.0	0.9975	3.42	0.60	10.8
37	8.1	0.380	0.28	2.1	NaN	13.0	30.0	0.9968	3.23	0.73	9.7
38	5.7	1.130	0.09	1.5	0.172	7.0	19.0	0.9940	3.50	0.48	9.8
39	7.3	0.450	0.36	5.9	0.074	12.0	87.0	0.9978	3.33	0.83	10.5
40	7.3	0.450	0.36	5.9	0.074	12.0	87.0	0.9978	3.33	0.83	10.5
41	8.8	0.610	0.30	2.8	0.088	17.0	46.0	0.9976	3.26	0.51	9.3
42	7.5	0.490	0.20	2.6	0.332	8.0	14.0	0.9968	3.21	0.90	10.5
43	8.1	0.660	0.22	2.2	0.069	9.0	23.0	0.9968	3.30	1.20	10.3
44	6.8	0.670	0.02	1.8	NaN	5.0	11.0	0.9962	3.48	0.52	9.5
45	4.6	0.520	0.15	2.1	0.054	8.0	65.0	0.9934	3.90	0.56	13.1
46	7.7	0.935	0.43	2.2	0.114	22.0	114.0	0.9970	3.25	0.73	9.2
47	8.7	0.290	0.52	1.6	0.113	12.0	37.0	0.9969	3.25	0.58	9.5
48	6.4	0.400	0.23	1.6	NaN	5.0	12.0	0.9958	3.34	0.56	9.2
49	5.6	0.310	0.37	1.4	0.074	12.0	96.0	0.9954	3.32	0.58	9.2

In [54]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   fixed acidity         50 non-null     float64
 1   volatile acidity      50 non-null     float64
 2   citric acid           50 non-null     float64
 3   residual sugar        50 non-null     float64
 4   chlorides             38 non-null     float64
 5   free sulfur dioxide   50 non-null     float64
 6   total sulfur dioxide  50 non-null     float64
 7   density               50 non-null     float64
 8   pH                    50 non-null     float64
 9   sulphates             50 non-null     float64
 10  alcohol               50 non-null     float64
 11  quality               50 non-null     int64  
```

```
dtypes: float64(11), int64(1)
memory usage: 4.8 KB
```

In [55]:

```
df1 = df.select_dtypes(include=[np.number])
df1['chlorides'] = df['chlorides'].fillna(median_h)
```

In [56]:

```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   fixed acidity         50 non-null     float64
 1   volatile acidity      50 non-null     float64
 2   citric acid           50 non-null     float64
 3   residual sugar        50 non-null     float64
 4   chlorides             50 non-null     float64
 5   free sulfur dioxide   50 non-null     float64
 6   total sulfur dioxide  50 non-null     float64
 7   density               50 non-null     float64
 8   pH                    50 non-null     float64
 9   sulphates             50 non-null     float64
 10  alcohol               50 non-null     float64
 11  quality               50 non-null     int64
dtypes: float64(11), int64(1)
memory usage: 4.8 KB
```

In [67]:

```
min = np.percentile(df1.sulphates, 5)
max = np.percentile(df1.sulphates, 95)
print(min, max)
```

```
0.489 1.1459999999999997
```

In [68]:

```
df1[(df1.sulphates < min)] = min
df1[(df1.sulphates > max)] = max
```

In [69]:

```
df1
```

Out[69]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol
0	7.400	0.700	0.000	1.900	0.076	11.000	34.000	0.9978	3.510	0.560	9.400
1	7.800	0.880	0.000	2.600	0.098	25.000	67.000	0.9968	3.200	0.680	9.800
2	7.800	0.760	0.040	2.300	0.086	15.000	54.000	0.9970	3.260	0.650	9.800
3	11.200	0.280	0.560	1.900	0.075	17.000	60.000	0.9980	3.160	0.580	9.800
4	7.400	0.700	0.000	1.900	0.086	11.000	34.000	0.9978	3.510	0.560	9.400
5	7.400	0.660	0.000	1.800	0.075	13.000	40.000	0.9978	3.510	0.560	9.400
6	0.489	0.489	0.489	0.489	0.489	0.489	0.489	0.4890	0.489	0.489	0.489
7	0.489	0.489	0.489	0.489	0.489	0.489	0.489	0.4890	0.489	0.489	0.489
8	7.800	0.580	0.020	2.000	0.086	9.000	18.000	0.9968	3.360	0.570	9.500
9	7.500	0.500	0.360	6.100	0.071	17.000	102.000	0.9978	3.350	0.800	10.500

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol
10	6.700	0.580	0.080	1.800	0.097	15.000	65.000	0.9959	3.280	0.540	9.200
11	7.500	0.500	0.360	6.100	0.086	17.000	102.000	0.9978	3.350	0.800	10.500
12	5.600	0.615	0.000	1.600	0.089	16.000	59.000	0.9943	3.580	0.520	9.900
13	1.146	1.146	1.146	1.146	1.146	1.146	1.146	1.1460	1.146	1.146	1.146
14	8.900	0.620	0.180	3.800	0.176	52.000	145.000	0.9986	3.160	0.880	9.200
15	8.900	0.620	0.190	3.900	0.170	51.000	148.000	0.9986	3.170	0.930	9.200
16	8.500	0.280	0.560	1.800	0.092	35.000	103.000	0.9969	3.300	0.750	10.500
17	1.146	1.146	1.146	1.146	1.146	1.146	1.146	1.1460	1.146	1.146	1.146
18	7.400	0.590	0.080	4.400	0.086	6.000	29.000	0.9974	3.380	0.500	9.000
19	7.900	0.320	0.510	1.800	0.341	17.000	56.000	0.9969	3.040	1.080	9.200
20	8.900	0.220	0.480	1.800	0.077	29.000	60.000	0.9968	3.390	0.530	9.400
21	7.600	0.390	0.310	2.300	0.082	23.000	71.000	0.9982	3.520	0.650	9.700
22	7.900	0.430	0.210	1.600	0.106	10.000	37.000	0.9966	3.170	0.910	9.500
23	8.500	0.490	0.110	2.300	0.086	9.000	67.000	0.9968	3.170	0.530	9.400
24	6.900	0.400	0.140	2.400	0.085	21.000	40.000	0.9968	3.430	0.630	9.700
25	6.300	0.390	0.160	1.400	0.086	11.000	23.000	0.9955	3.340	0.560	9.300
26	7.600	0.410	0.240	1.800	0.080	4.000	11.000	0.9962	3.280	0.590	9.500
27	7.900	0.430	0.210	1.600	0.106	10.000	37.000	0.9966	3.170	0.910	9.500
28	7.100	0.710	0.000	1.900	0.080	14.000	35.000	0.9972	3.470	0.550	9.400
29	7.800	0.645	0.000	2.000	0.086	8.000	16.000	0.9964	3.380	0.590	9.800
30	6.700	0.675	0.070	2.400	0.089	17.000	82.000	0.9958	3.350	0.540	10.100
31	6.900	0.685	0.000	2.500	0.105	22.000	37.000	0.9966	3.460	0.570	10.600
32	8.300	0.655	0.120	2.300	0.083	15.000	113.000	0.9966	3.170	0.660	9.800
33	6.900	0.605	0.120	10.700	0.073	40.000	83.000	0.9993	3.450	0.520	9.400
34	5.200	0.320	0.250	1.800	0.086	13.000	50.000	0.9957	3.380	0.550	9.200
35	7.800	0.645	0.000	5.500	0.086	5.000	18.000	0.9986	3.400	0.550	9.600
36	7.800	0.600	0.140	2.400	0.086	3.000	15.000	0.9975	3.420	0.600	10.800
37	8.100	0.380	0.280	2.100	0.086	13.000	30.000	0.9968	3.230	0.730	9.700
38	0.489	0.489	0.489	0.489	0.489	0.489	0.489	0.4890	0.489	0.489	0.489
39	7.300	0.450	0.360	5.900	0.074	12.000	87.000	0.9978	3.330	0.830	10.500
40	7.300	0.450	0.360	5.900	0.074	12.000	87.000	0.9978	3.330	0.830	10.500
41	8.800	0.610	0.300	2.800	0.088	17.000	46.000	0.9976	3.260	0.510	9.300
42	7.500	0.490	0.200	2.600	0.332	8.000	14.000	0.9968	3.210	0.900	10.500
43	1.146	1.146	1.146	1.146	1.146	1.146	1.146	1.1460	1.146	1.146	1.146
44	6.800	0.670	0.020	1.800	0.086	5.000	11.000	0.9962	3.480	0.520	9.500

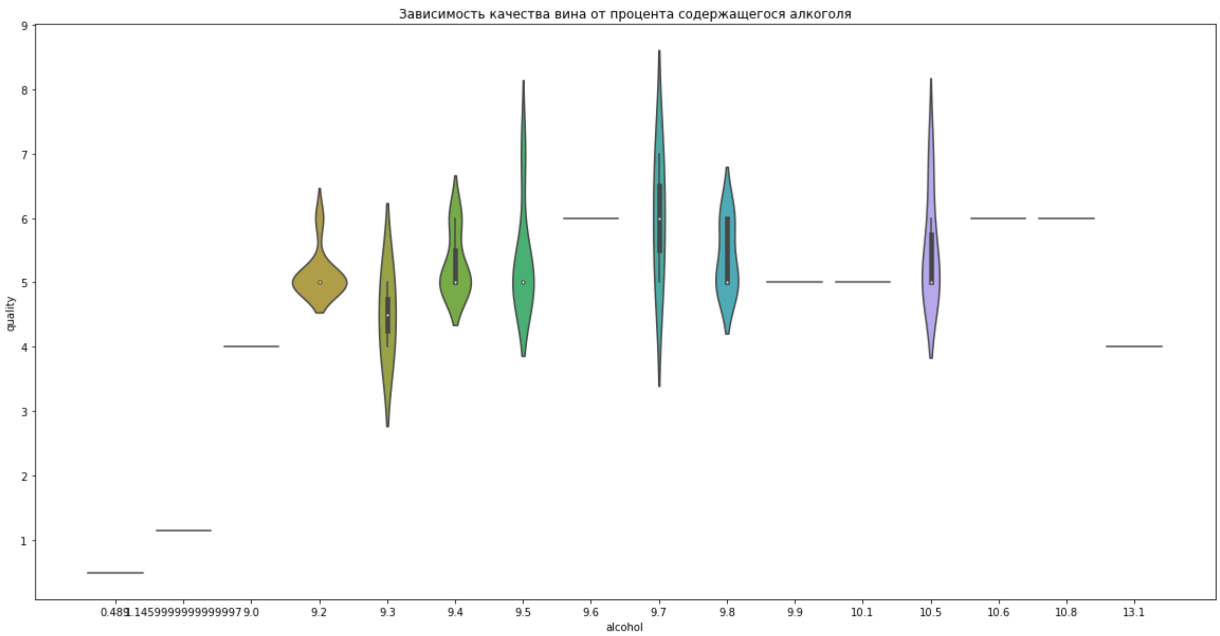
	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol
45	4.600	0.520	0.150	2.100	0.054	8.000	65.000	0.9934	3.900	0.560	13.100
46	7.700	0.935	0.430	2.200	0.114	22.000	114.000	0.9970	3.250	0.730	9.200
47	8.700	0.290	0.520	1.600	0.113	12.000	37.000	0.9969	3.250	0.580	9.500
48	6.400	0.400	0.230	1.600	0.086	5.000	12.000	0.9958	3.340	0.560	9.200
49	5.600	0.310	0.370	1.400	0.074	12.000	96.000	0.9954	3.320	0.580	9.200

In [87]:

```
plt.figure(figsize=(20, 10))
ax = sns.violinplot(df1.alcohol, df1.quality)
ax.set(xlabel='alcohol', ylabel='quality', title='Зависимость качества вина от проце
plt.plot()
```

```
c:\users\sveta\documents\virtualenvs\tensorflow\lib\site-packages\seaborn\_decorator
s.py:43: FutureWarning: Pass the following variables as keyword args: x, y. From ver
sion 0.12, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretation.
  FutureWarning
```

Out[87]: []



In [ ]: