

In [1]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import seaborn as pairplot
import warnings
warnings.filterwarnings('ignore')
```

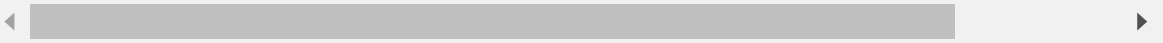
In [2]:

```
df=pd.read_csv('wine.csv')
df
```

Out[2]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates
0	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	0.56
1	7.8	0.880	0.00	2.6	0.098	25.0	67.0	0.99680	3.20	0.68
2	7.8	0.760	0.04	2.3	0.092	15.0	54.0	0.99700	3.26	0.65
3	11.2	0.280	0.56	1.9	0.075	17.0	60.0	0.99800	3.16	0.58
4	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	0.56
...	...	...	...	...	...	...	...	...	...	...
1594	6.2	0.600	0.08	2.0	0.090	32.0	44.0	0.99490	3.45	0.58
1595	5.9	0.550	0.10	2.2	0.062	39.0	51.0	0.99512	3.52	0.76
1596	6.3	0.510	0.13	2.3	0.076	29.0	40.0	0.99574	3.42	0.75
1597	5.9	0.645	0.12	2.0	0.075	32.0	44.0	0.99547	3.57	0.71
1598	6.0	0.310	0.47	3.6	0.067	18.0	42.0	0.99549	3.39	0.66

1599 rows × 12 columns



In [3]:

```
df.head()
```

Out[3]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alc
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	

In [4]:

```
df.tail()
```

Out[4]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alc
1594	6.2	0.600	0.08	2.0	0.090	32.0	44.0	0.99490	3.45	0.58	
1595	5.9	0.550	0.10	2.2	0.062	39.0	51.0	0.99512	3.52	0.76	
1596	6.3	0.510	0.13	2.3	0.076	29.0	40.0	0.99574	3.42	0.75	
1597	5.9	0.645	0.12	2.0	0.075	32.0	44.0	0.99547	3.57	0.71	
1598	6.0	0.310	0.47	3.6	0.067	18.0	42.0	0.99549	3.39	0.66	

In [5]:

```
df.sample(3)
```

Out[5]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alc
1505	6.7	0.76	0.02	1.8	0.078	6.0	12.0	0.99600	3.55	0.63	
1044	6.4	0.39	0.33	3.3	0.046	12.0	53.0	0.99294	3.36	0.62	
911	9.1	0.28	0.46	9.0	0.114	3.0	9.0	0.99901	3.18	0.60	

In [6]:

```
df.shape
```

Out[6]:

(1599, 12)

In [7]:

```
df.describe()
```

Out[7]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide
count	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000
mean	8.319637	0.527821	0.270976	2.538806	0.087467	15.874922	46.206916
std	1.741096	0.179060	0.194801	1.409928	0.047065	10.460157	32.041722
min	4.600000	0.120000	0.000000	0.900000	0.012000	1.000000	6.010000
25%	7.100000	0.390000	0.090000	1.900000	0.070000	7.000000	22.010000
50%	7.900000	0.520000	0.260000	2.200000	0.079000	14.000000	38.010000
75%	9.200000	0.640000	0.420000	2.600000	0.090000	21.000000	62.010000
max	15.900000	1.580000	1.000000	15.500000	0.611000	72.000000	289.000000

In [9]:

```
df.dtypes
```

Out[9]:

```
fixed acidity      float64
volatile acidity   float64
citric acid        float64
residual sugar     float64
chlorides          float64
free sulfur dioxide float64
total sulfur dioxide float64
density            float64
pH                float64
sulphates          float64
alcohol            float64
quality            int64
dtype: object
```

In [10]:

```
df.columns
```

Out[10]:

```
Index(['fixed acidity', 'volatile acidity', 'citric acid', 'residual suga
r',
      'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'densit
y',
      'pH', 'sulphates', 'alcohol', 'quality'],
      dtype='object')
```

In [11]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1599 entries, 0 to 1598
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   fixed acidity          1599 non-null   float64
1   volatile acidity       1599 non-null   float64
2   citric acid            1599 non-null   float64
3   residual sugar         1599 non-null   float64
4   chlorides              1599 non-null   float64
5   free sulfur dioxide    1599 non-null   float64
6   total sulfur dioxide   1599 non-null   float64
7   density               1599 non-null   float64
8   pH                    1599 non-null   float64
9   sulphates             1599 non-null   float64
10  alcohol               1599 non-null   float64
11  quality               1599 non-null   int64
dtypes: float64(11), int64(1)
memory usage: 150.0 KB
```

In [12]:

```
df['quality'].unique()
```

Out[12]:

```
array([5, 6, 7, 4, 8, 3], dtype=int64)
```

In [13]:

```
quality_ids={5:0,6:1,7:2,4:3,8:4,3:5}
```

In [15]:

```
df['quality_ids']=df.quality.map(quality_ids)
```

In [16]:

df

Out[16]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates
0	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	0.56
1	7.8	0.880	0.00	2.6	0.098	25.0	67.0	0.99680	3.20	0.68
2	7.8	0.760	0.04	2.3	0.092	15.0	54.0	0.99700	3.26	0.65
3	11.2	0.280	0.56	1.9	0.075	17.0	60.0	0.99800	3.16	0.58
4	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	0.56
...	...	...	...	...	...	...	...	...	...	...
1594	6.2	0.600	0.08	2.0	0.090	32.0	44.0	0.99490	3.45	0.58
1595	5.9	0.550	0.10	2.2	0.062	39.0	51.0	0.99512	3.52	0.76
1596	6.3	0.510	0.13	2.3	0.076	29.0	40.0	0.99574	3.42	0.75
1597	5.9	0.645	0.12	2.0	0.075	32.0	44.0	0.99547	3.57	0.71
1598	6.0	0.310	0.47	3.6	0.067	18.0	42.0	0.99549	3.39	0.66

1599 rows × 13 columns

In [51]:

df['quality'].value\_counts().sum()

Out[51]:

1599

In [52]:

df['quality'].value\_counts().sort\_index()

Out[52]:

```

3      10
4       53
5     681
6     638
7     199
8       18
Name: quality, dtype: int64

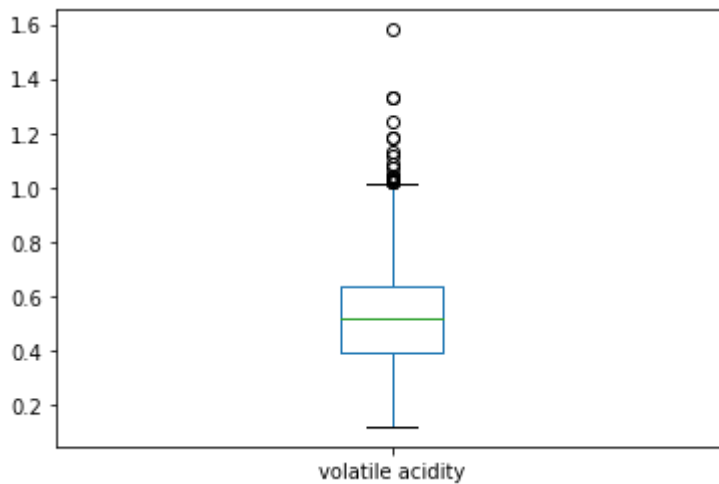
```

In [18]:

```
df['volatile acidity'].plot.box()
```

Out[18]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x13ca8e96108>

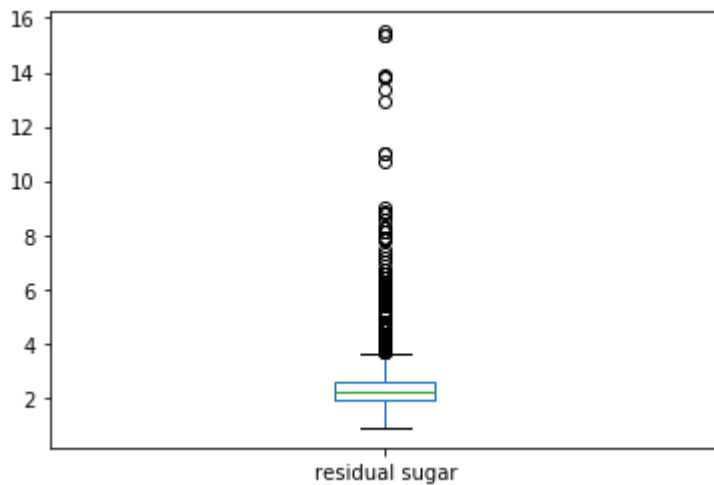


In [19]:

```
df['residual sugar'].plot.box()
```

Out[19]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x13ca74e5888>

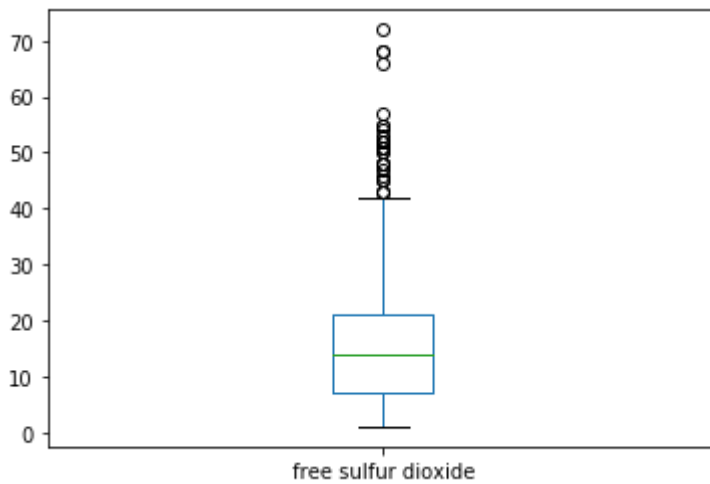


In [20]:

```
df['free sulfur dioxide'].plot.box()
```

Out[20]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x13ca96a71c8>

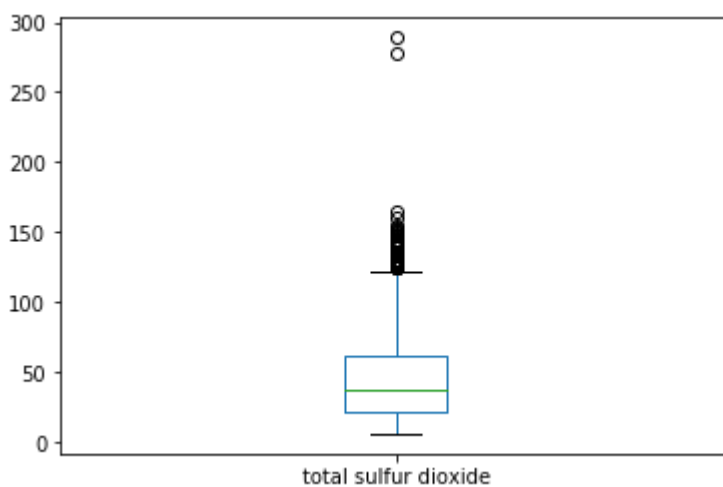


In [21]:

```
df['total sulfur dioxide'].plot.box()
```

Out[21]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x13ca9714048>



In [22]:

```
dfcorr=df.corr()  
dfcorr
```

Out[22]:

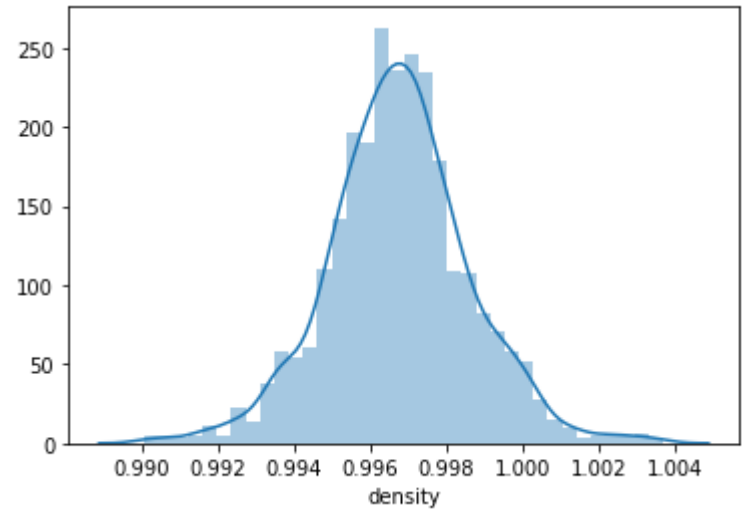
	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	der
fixed acidity	1.000000	-0.256131	0.671703	0.114777	0.093705	-0.153794	-0.113181	0.668
volatile acidity	-0.256131	1.000000	-0.552496	0.001918	0.061298	-0.010504	0.076470	0.022
citric acid	0.671703	-0.552496	1.000000	0.143577	0.203823	-0.060978	0.035533	0.364
residual sugar	0.114777	0.001918	0.143577	1.000000	0.055610	0.187049	0.203028	0.355
chlorides	0.093705	0.061298	0.203823	0.055610	1.000000	0.005562	0.047400	0.200
free sulfur dioxide	-0.153794	-0.010504	-0.060978	0.187049	0.005562	1.000000	0.667666	-0.021
total sulfur dioxide	-0.113181	0.076470	0.035533	0.203028	0.047400	0.667666	1.000000	0.071
density	0.668047	0.022026	0.364947	0.355283	0.200632	-0.021946	0.071269	1.000
pH	-0.682978	0.234937	-0.541904	-0.085652	-0.265026	0.070377	-0.066495	-0.341
sulphates	0.183006	-0.260987	0.312770	0.005527	0.371260	0.051658	0.042947	0.148
alcohol	-0.061668	-0.202288	0.109903	0.042075	-0.221141	-0.069408	-0.205654	-0.496
quality	0.124052	-0.390558	0.226373	0.013732	-0.128907	-0.050656	-0.185100	-0.174
quality_ids	0.066688	-0.120807	0.109674	0.031412	-0.072503	-0.118959	-0.242310	-0.160

In [23]:

```
sns.distplot(df['density'])
```

Out[23]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x13ca9792208>



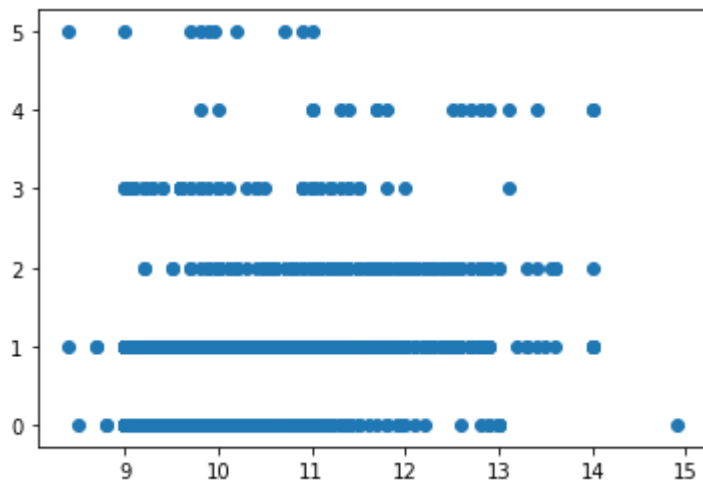


In [24]:

```
plt.scatter(df['alcohol'],df['quality_ids'])
```

Out[24]:

<matplotlib.collections.PathCollection at 0x13ca975ce08>

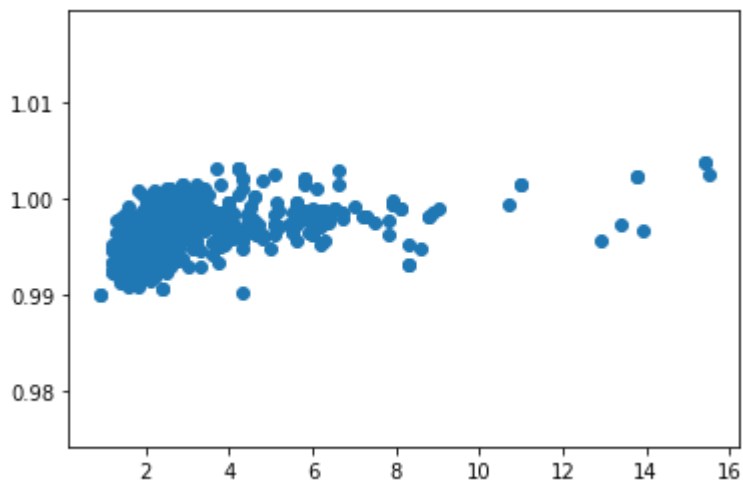


In [25]:

```
plt.scatter(df['residual sugar'],df['density'])
```

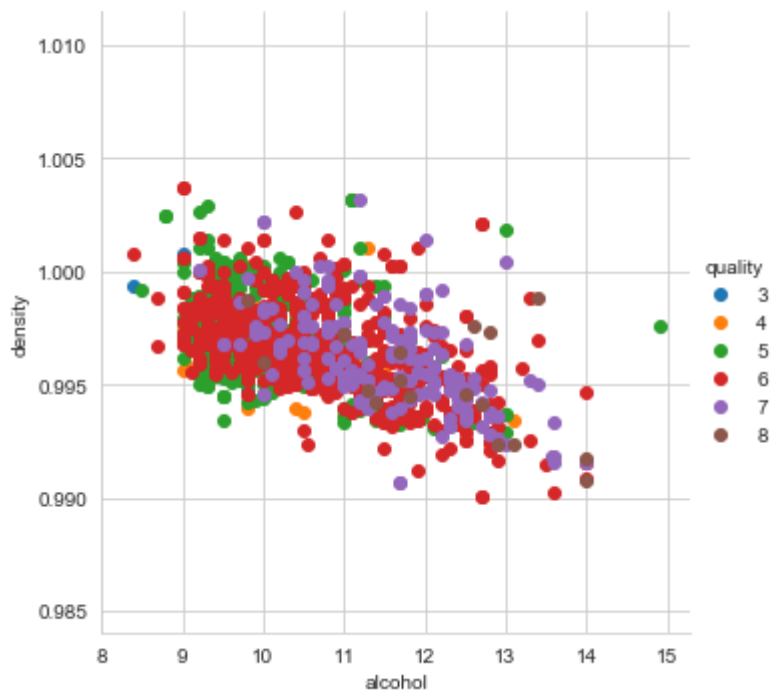
Out[25]:

<matplotlib.collections.PathCollection at 0x13ca9900d48>



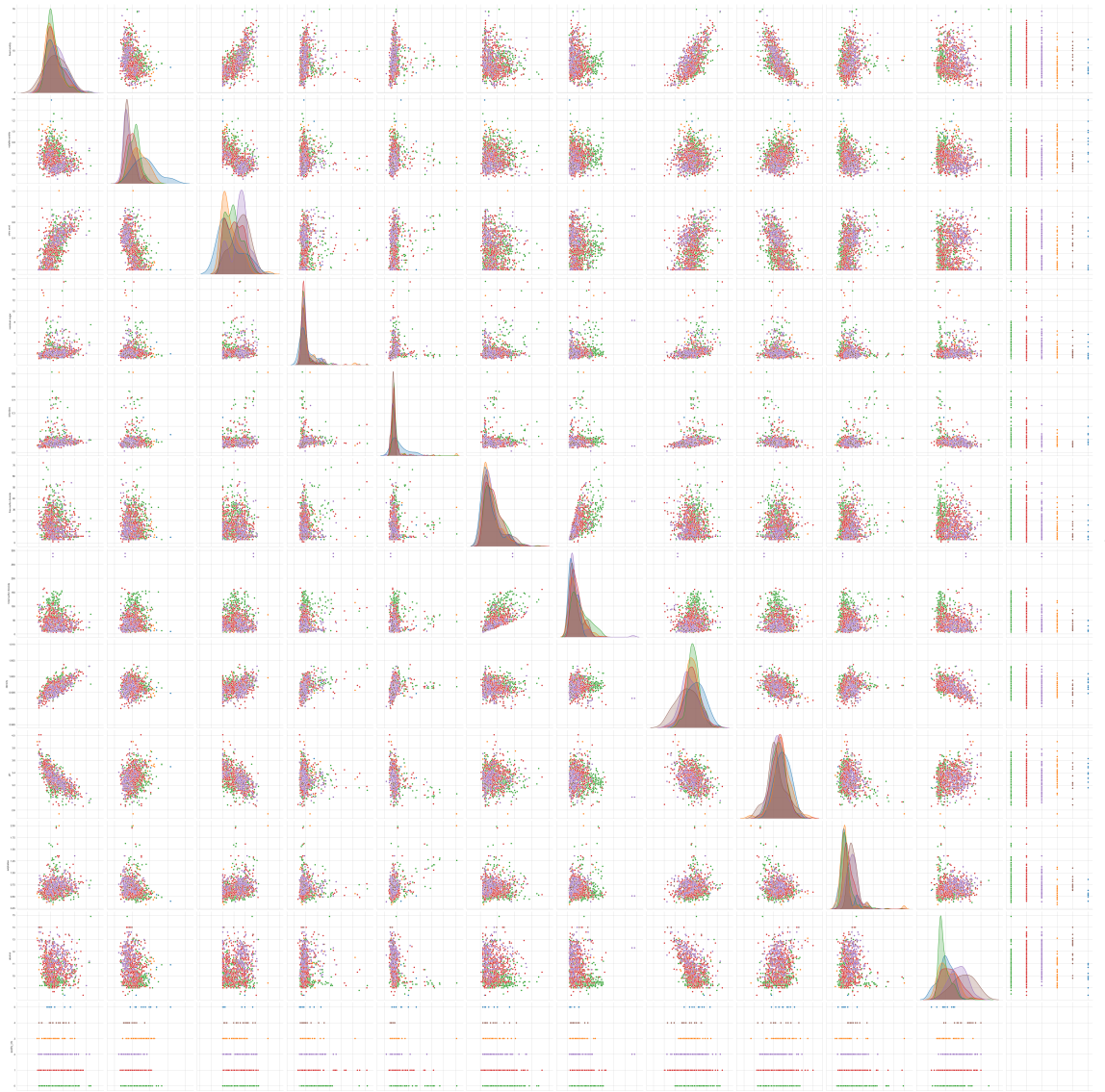
In [45]:

```
sns.set_style("whitegrid");  
sns.FacetGrid(df, hue='quality', size=5)\  
    .map(plt.scatter, 'alcohol', 'density')\  
    .add_legend();  
plt.show()
```



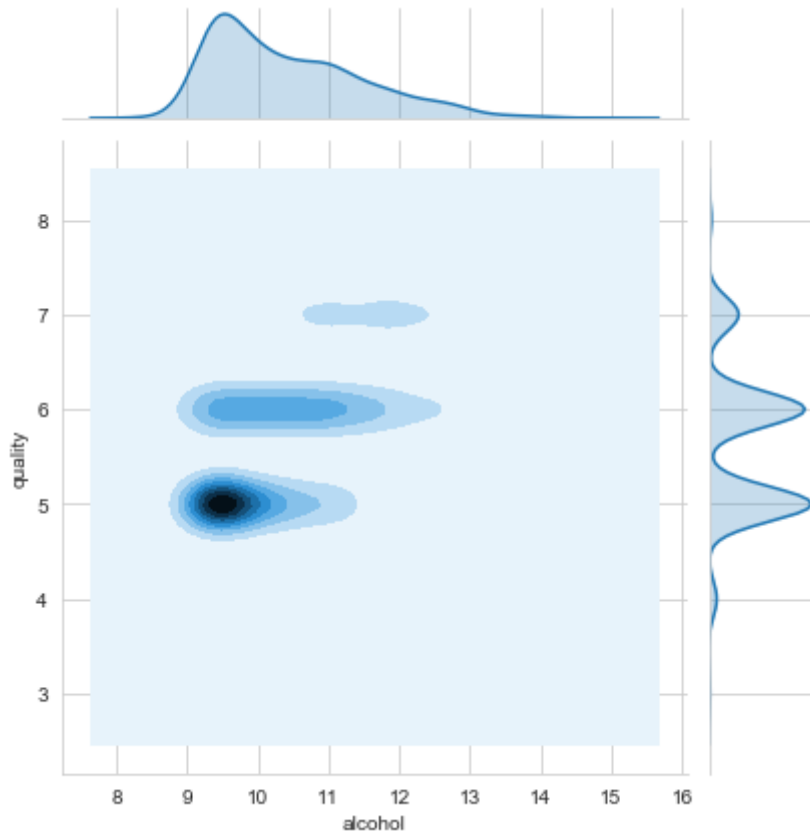
In [49]:

```
plt.close()
sns.set_style("whitegrid");
sns.pairplot(df,hue='quality',size=5);
plt.show()
```



In [50]:

```
sns.jointplot(x='alcohol',y='quality',kind="kde",data=df)  
plt.show()
```



In [ ]: