In [31]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [32]:
```
df=pd.read_csv(r"C:\Users\user\Downloads\11_winequality-red.csv")
df.fillna(0,inplace=True)
df
```

Out[32]:

| | fixed acidity | volatile acidity | citric acid | residual sugar | chlorides | free sulfur dioxide | total sulfur dioxide | density | pH | sulphates | alcol |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **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 | 1 |
| **1595** | 5.9 | 0.550 | 0.10 | 2.2 | 0.062 | 39.0 | 51.0 | 0.99512 | 3.52 | 0.76 | 1 |
| **1596** | 6.3 | 0.510 | 0.13 | 2.3 | 0.076 | 29.0 | 40.0 | 0.99574 | 3.42 | 0.75 | 1 |
| **1597** | 5.9 | 0.645 | 0.12 | 2.0 | 0.075 | 32.0 | 44.0 | 0.99547 | 3.57 | 0.71 | 1 |
| **1598** | 6.0 | 0.310 | 0.47 | 3.6 | 0.067 | 18.0 | 42.0 | 0.99549 | 3.39 | 0.66 | 1 |

1599 rows × 12 columns

In [33]:
```
df.head()
```

Out[33]:

| | fixed acidity | volatile acidity | citric acid | residual sugar | chlorides | free sulfur dioxide | total sulfur dioxide | density | pH | sulphates | alcohol |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 7.4 | 0.70 | 0.00 | 1.9 | 0.076 | 11.0 | 34.0 | 0.9978 | 3.51 | 0.56 | 9.4 |
| **1** | 7.8 | 0.88 | 0.00 | 2.6 | 0.098 | 25.0 | 67.0 | 0.9968 | 3.20 | 0.68 | 9.8 |
| **2** | 7.8 | 0.76 | 0.04 | 2.3 | 0.092 | 15.0 | 54.0 | 0.9970 | 3.26 | 0.65 | 9.8 |
| **3** | 11.2 | 0.28 | 0.56 | 1.9 | 0.075 | 17.0 | 60.0 | 0.9980 | 3.16 | 0.58 | 9.8 |
| **4** | 7.4 | 0.70 | 0.00 | 1.9 | 0.076 | 11.0 | 34.0 | 0.9978 | 3.51 | 0.56 | 9.4 |

In [34]:
```
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 [35]:
```
import seaborn as sns
```

In [36]:
```
df.describe()
```

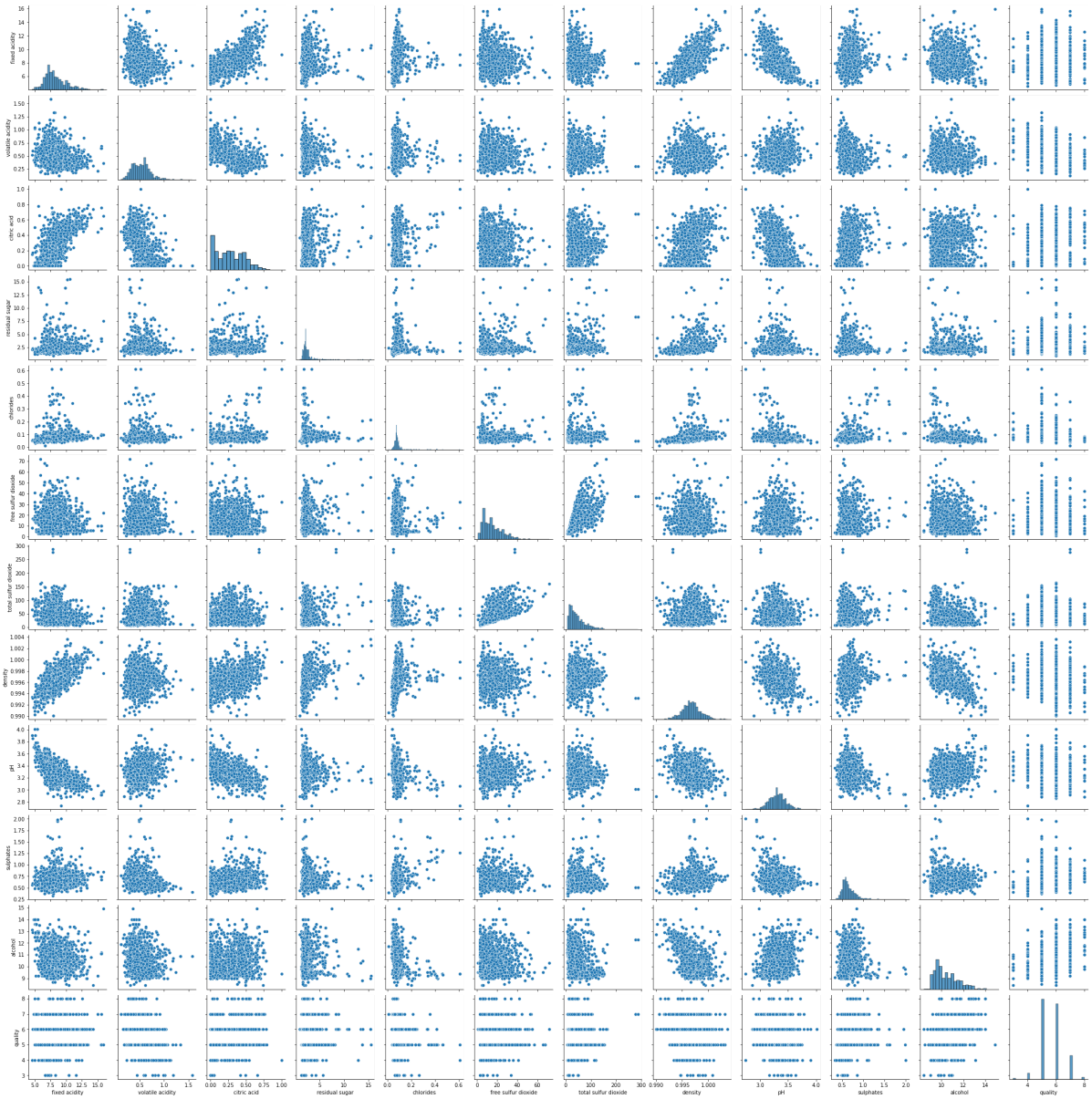
Out[36]:

| | fixed acidity | volatile acidity | citric acid | residual sugar | chlorides | free sulfur dioxide | total sulfu dioxid |
|---|---|---|---|---|---|---|---|
| **count** | 1599.000000 | 1599.000000 | 1599.000000 | 1599.000000 | 1599.000000 | 1599.000000 | 1599.00000 |
| **mean** | 8.319637 | 0.527821 | 0.270976 | 2.538806 | 0.087467 | 15.874922 | 46.46779 |
| **std** | 1.741096 | 0.179060 | 0.194801 | 1.409928 | 0.047065 | 10.460157 | 32.89532 |
| **min** | 4.600000 | 0.120000 | 0.000000 | 0.900000 | 0.012000 | 1.000000 | 6.00000 |
| **25%** | 7.100000 | 0.390000 | 0.090000 | 1.900000 | 0.070000 | 7.000000 | 22.00000 |
| **50%** | 7.900000 | 0.520000 | 0.260000 | 2.200000 | 0.079000 | 14.000000 | 38.00000 |
| **75%** | 9.200000 | 0.640000 | 0.420000 | 2.600000 | 0.090000 | 21.000000 | 62.00000 |
| **max** | 15.900000 | 1.580000 | 1.000000 | 15.500000 | 0.611000 | 72.000000 | 289.00000 |

In [10]:
```

```

In [37]:
```
sns.pairplot(df)
```

Out[37]:
```
<seaborn.axisgrid.PairGrid at 0x1c357211940>
```

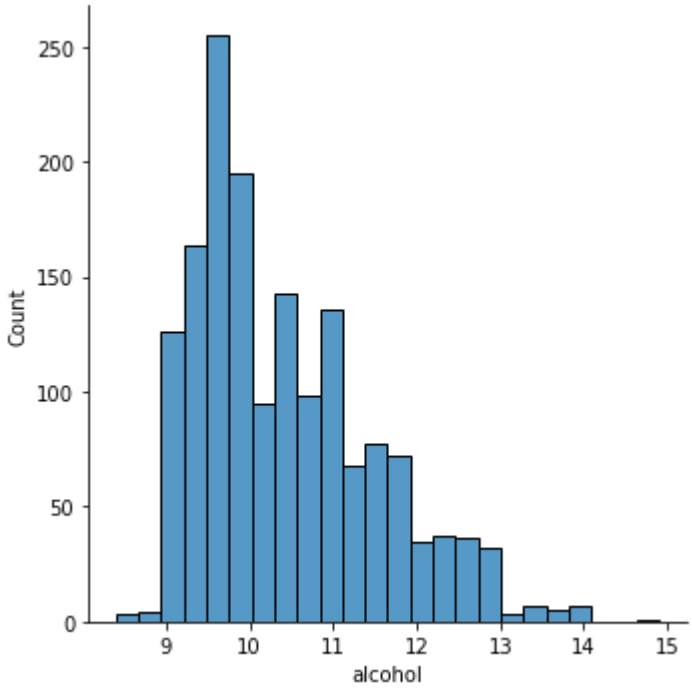


In [38]:
```
df1=df.drop(['citric acid'],axis=1)
df1
df1=df1.drop(df1.index[1537:])
df1.isna().sum()
```

Out[38]:
```
fixed acidity           0
volatile acidity        0
residual sugar          0
chlorides               0
free sulfur dioxide     0
total sulfur dioxide    0
density                 0
pH                      0
sulphates               0
alcohol                 0
quality                 0
dtype: int64
```

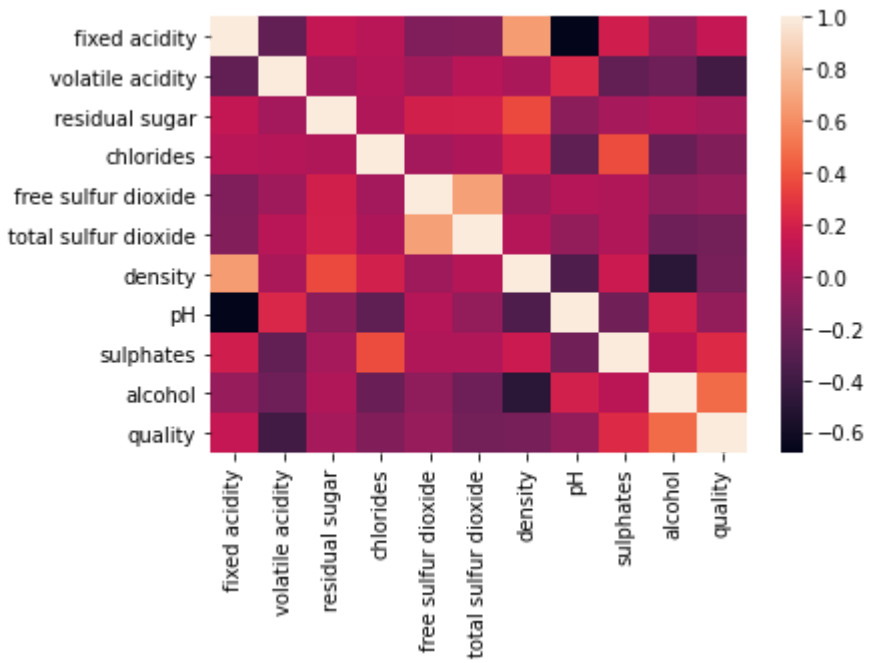
In [39]:
```
sns.displot(df['alcohol'])
```

Out[39]:
```
<seaborn.axisgrid.FacetGrid at 0x1c3533a32b0>
```



In [40]:
```
sns.heatmap(df1.corr())
```

Out[40]: <AxesSubplot:>




In [41]:
```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

In [42]:
```
df1.isna().sum()
```

Out[42]:
```
fixed acidity           0
volatile acidity        0
residual sugar          0
chlorides               0
free sulfur dioxide     0
total sulfur dioxide    0
density                 0
pH                      0
sulphates               0
alcohol                 0
quality                 0
dtype: int64
```

In [44]:
```
y=df1['fixed acidity']
x=df1.drop(['chlorides','residual sugar'],axis=1)
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3)
print(x_train)
```

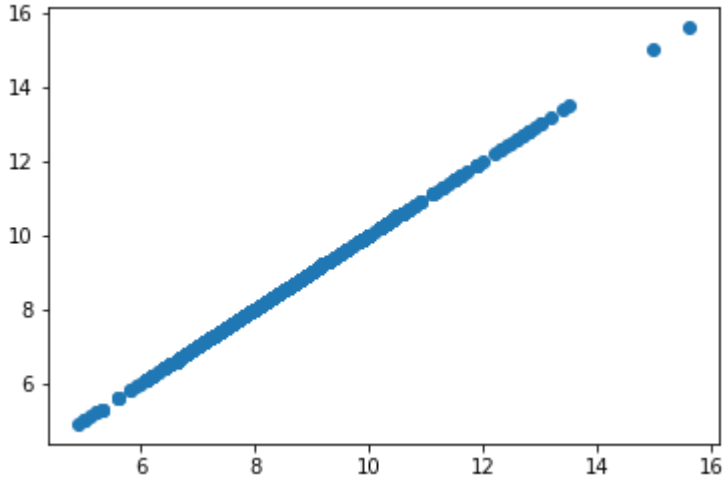
```
      fixed acidity  volatile acidity  free sulfur dioxide  \
713             8.0             0.430                 10.0
891             7.2             0.660                 16.0
25              6.3             0.390                 11.0
730             9.5             0.550                 12.0
606             9.4             0.410                 10.0
...             ...               ...                  ...
1055            8.2             0.640                  5.0
1272            5.9             0.460                 25.0
1096            6.6             0.725                  9.0
1334            7.2             0.835                  4.0
1333            9.1             0.775                 12.0

      total sulfur dioxide  density    pH  sulphates  alcohol  quality
713                   48.0  0.99760  3.34       0.46      9.4        5
891                   86.0  0.99743  3.53       0.57      9.7        5
25                    23.0  0.99550  3.34       0.56      9.3        5
730                   37.0  0.99820  3.17       0.67      9.6        5
606                   20.0  0.99730  3.34       0.79     12.2        7
...                    ...      ...   ...        ...      ...      ...
1055                  77.0  0.99747  3.13       0.62      9.1        6
1272                  44.0  0.99385  3.50       0.53     11.2        5
1096                  17.0  0.99655  3.35       0.49     10.8        6
1334                  11.0  0.99608  3.39       0.52     10.0        5
1333                  48.0  0.99760  3.18       0.51      9.6        5

[1075 rows x 9 columns]
```

In [45]:
```
model=LinearRegression()
model.fit(x_train,y_train)
model.intercept_
```

Out[45]:
```
5.1514348342607263e-14
```

In [46]:
```
prediction=model.predict(x_test)
plt.scatter(y_test,prediction)
```

Out[46]:
```
<matplotlib.collections.PathCollection at 0x1c35fc941f0>
```




In [47]:
```
model.score(x_test,y_test)
```

Out[47]:
```
1.0
```

In [48]:
```
from sklearn.linear_model import Ridge,Lasso
```

In [49]:
```
rr=Ridge(alpha=10)
rr.fit(x_train,y_train)
```

Out[49]:
```
Ridge(alpha=10)
```

In [50]:
```
rr.score(x_test,y_test)
```

Out[50]:
```
0.9999866935956242
```

In [51]:
```
la =Lasso(alpha=10)
la.fit(x_train,y_train)
```

Out[51]:
```
Lasso(alpha=10)
```

In [52]:
```
la.score(x_test,y_test)
```

Out[52]:
```
-0.0026296681839950153
```

In [ ]: