In [2]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings('ignore')
```

# Loading the Data from the CSV file

In [23]:

```
mData = pd.read_csv('Data Sets/Red wine quality/winequality-red.csv')
#printing the shape of the dataset
print('The Shape of The Data ',mData.shape)
```

```
The Shape of The Data  (1599, 12)
```

# Data Preprocessing

In [24]:

```
# getting some insights about the null values
mData["quality"] = mData["quality"].astype('object')
mData.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1599 entries, 0 to 1598
Data columns (total 12 columns):
fixed acidity           1599 non-null float64
volatile acidity        1599 non-null float64
citric acid             1599 non-null float64
residual sugar          1599 non-null float64
chlorides               1599 non-null float64
free sulfur dioxide     1599 non-null float64
total sulfur dioxide    1599 non-null float64
density                 1599 non-null float64
pH                      1599 non-null float64
sulphates               1599 non-null float64
alcohol                 1599 non-null float64
quality                 1599 non-null object
dtypes: float64(11), object(1)
memory usage: 150.0+ KB
```

In [25]:

```
Y = mData.quality
X = mData.drop(columns=['quality'])
X.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1599 entries, 0 to 1598
Data columns (total 11 columns):
fixed acidity           1599 non-null float64
volatile acidity        1599 non-null float64
citric acid             1599 non-null float64
residual sugar          1599 non-null float64
chlorides               1599 non-null float64
free sulfur dioxide     1599 non-null float64
total sulfur dioxide    1599 non-null float64
density                 1599 non-null float64
pH                      1599 non-null float64
sulphates               1599 non-null float64
alcohol                 1599 non-null float64
dtypes: float64(11)
memory usage: 137.5 KB
```

# Data Visualization

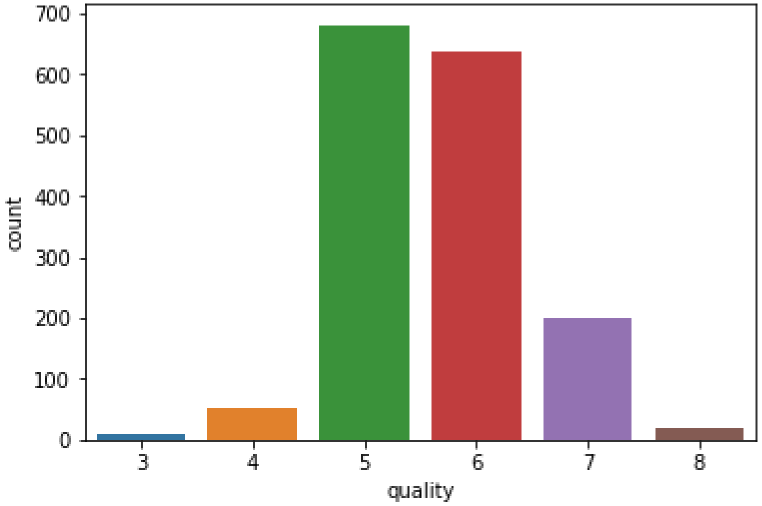
In [26]:

```
# Target Distribution
print(Y.value_counts())
sns.countplot(Y)
```

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

Out[26]:

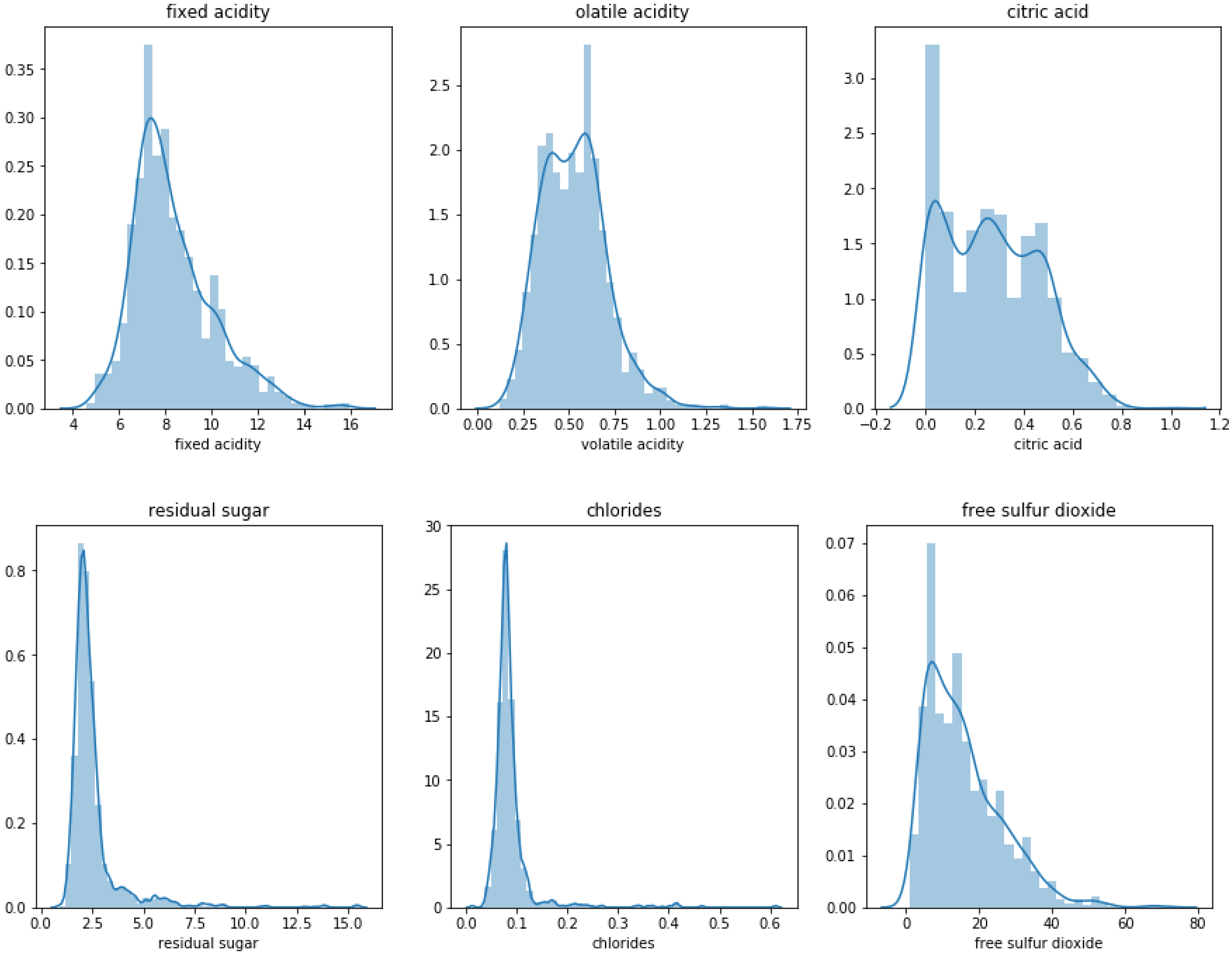
```
<matplotlib.axes._subplots.AxesSubplot at 0x11375bcf8>
```

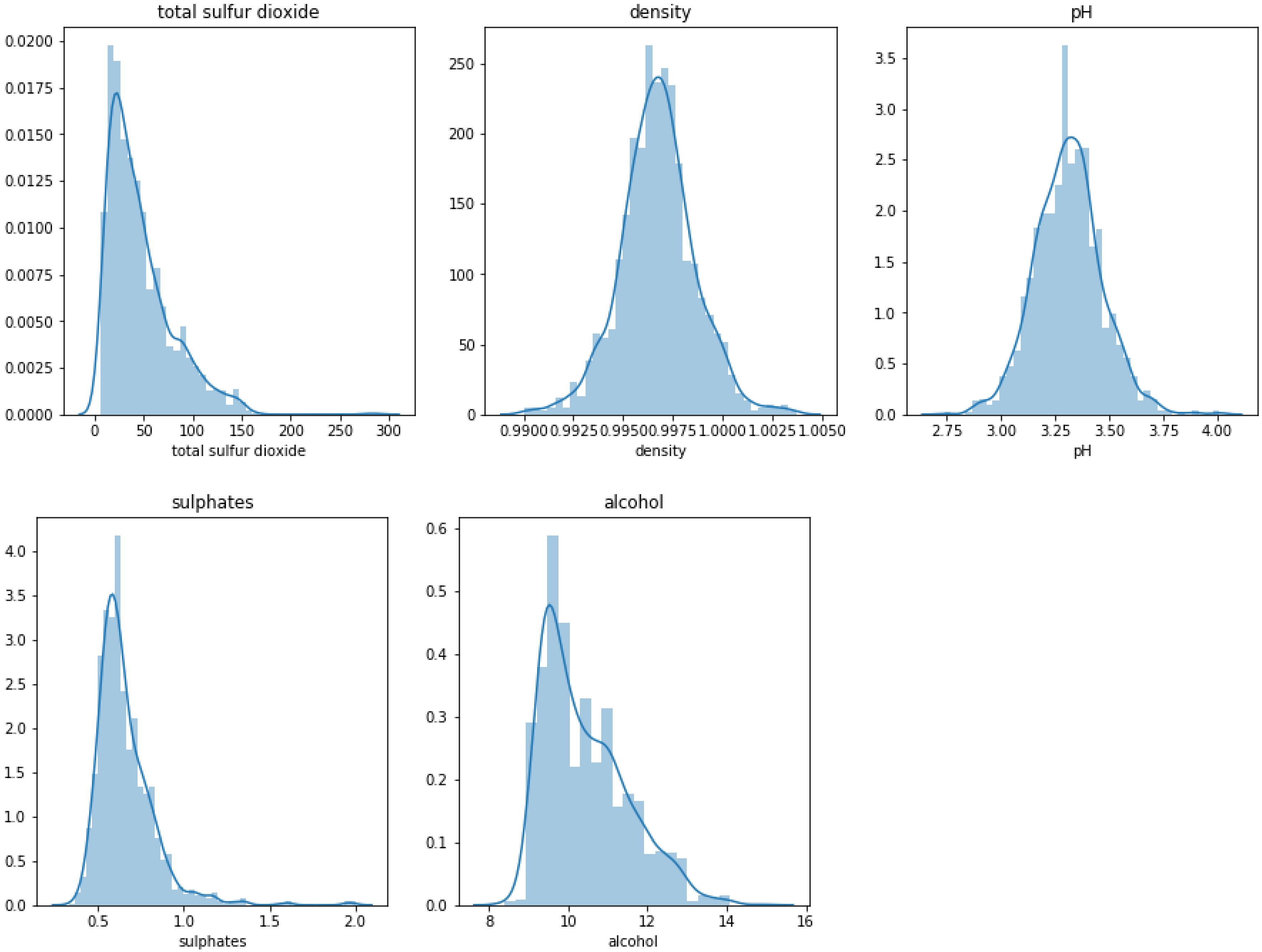


In [27]:

```
#histogram for all features
plt.figure(1,figsize=[15,5])
plt.subplot(1,3,1)
sns.distplot(X['fixed acidity'])
plt.title("fixed acidity")


plt.figure(1,figsize=[15,5])
plt.subplot(1,3,2)
sns.distplot(X['volatile acidity'])
plt.title("olatile acidity")

plt.figure(1,figsize=[15,5])
plt.subplot(1,3,3)
sns.distplot(X['citric acid'])
plt.title("citric acid")

plt.figure(2,figsize=[15,5])
plt.subplot(1,3,1)
sns.distplot(X['residual sugar'])
plt.title("residual sugar")


plt.figure(2,figsize=[15,5])
plt.subplot(1,3,2)
sns.distplot(X['chlorides'])
plt.title("chlorides")


plt.figure(2,figsize=[15,5])
plt.subplot(1,3,3)
sns.distplot(X['free sulfur dioxide'])
plt.title("free sulfur dioxide")

plt.figure(3,figsize=[15,5])
plt.subplot(1,3,1)
sns.distplot(X['total sulfur dioxide'])
plt.title("total sulfur dioxide")


plt.figure(3,figsize=[15,5])
plt.subplot(1,3,2)
sns.distplot(X['density'])
plt.title("density")


plt.figure(3,figsize=[15,5])
plt.subplot(1,3,3)
sns.distplot(X['pH'])
plt.title("pH")

plt.figure(4,figsize=[15,5])
plt.subplot(1,3,1)
sns.distplot(X['sulphates'])
plt.title("sulphates")

plt.figure(4,figsize=[15,5])
plt.subplot(1,3,2)
sns.distplot(X["alcohol"])
plt.title("alcohol")
```

```
Out[27]: Text(0.5,1,'alcohol')
```


fixed acidity
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
4
6
8
10
12
14
16
fixed acidity
olatile acidity
2.5
2.0
1.5
1.0
0.5
0.0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
volatile acidity
citric acid
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
citric acid
residual sugar
0.8
0.6
0.4
0.2
0.0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
residual sugar
chlorides
30
25
20
15
10
5
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
chlorides
free sulfur dioxide
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
20
40
60
80
free sulfur dioxide

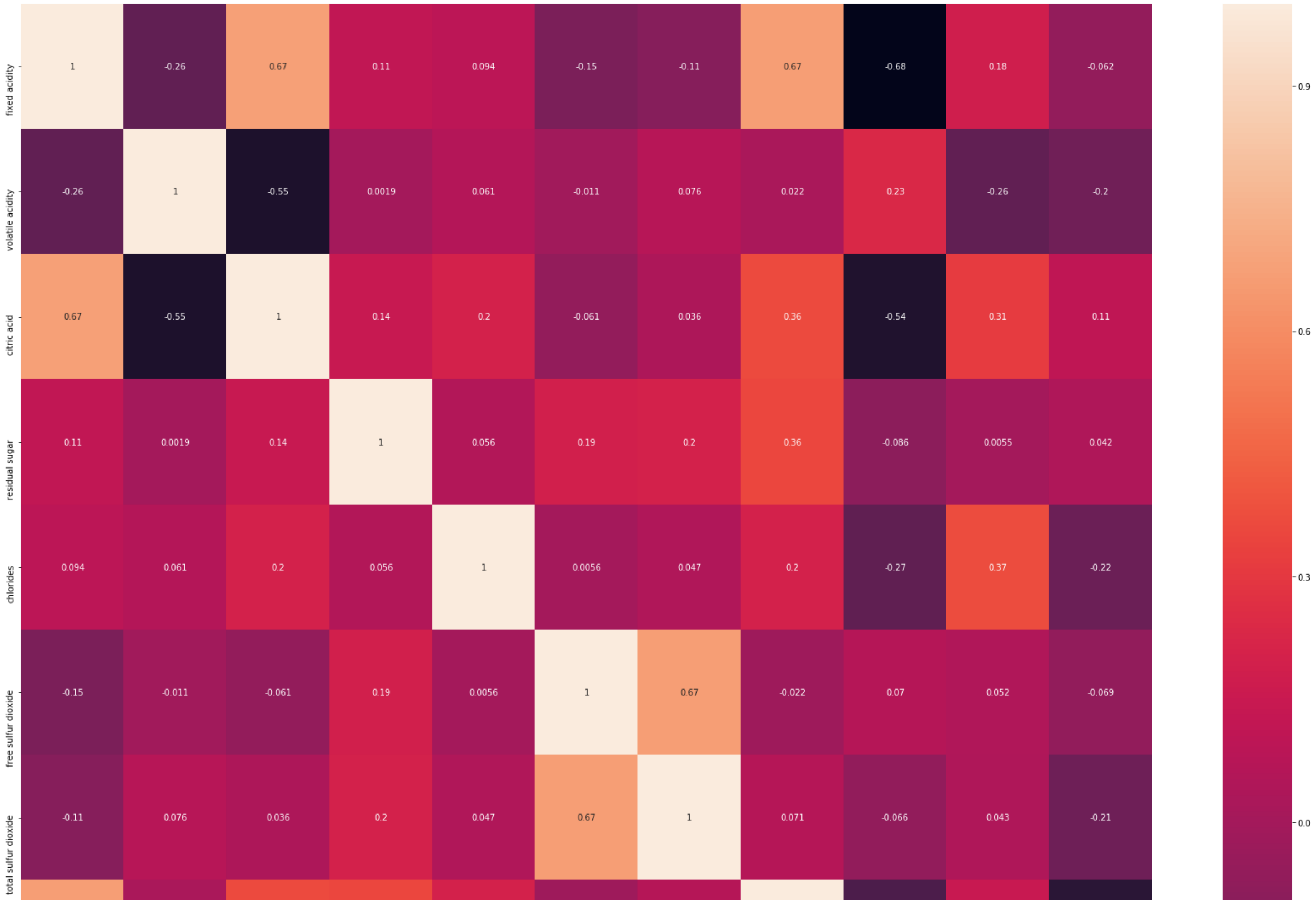

total sulfur dioxide
0.0200
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
0
50
100
150
200
250
300
total sulfur dioxide
density
250
200
150
100
50
0
0.9900
0.9925
0.9950
0.9975
1.0000
1.0025
1.0050
density
pH
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
2.75
3.00
3.25
3.50
3.75
4.00
pH
sulphates
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.5
1.0
1.5
2.0
sulphates
alcohol
0.6
0.5
0.4
0.3
0.2
0.1
0.0
8
10
12
14
16
alcohol

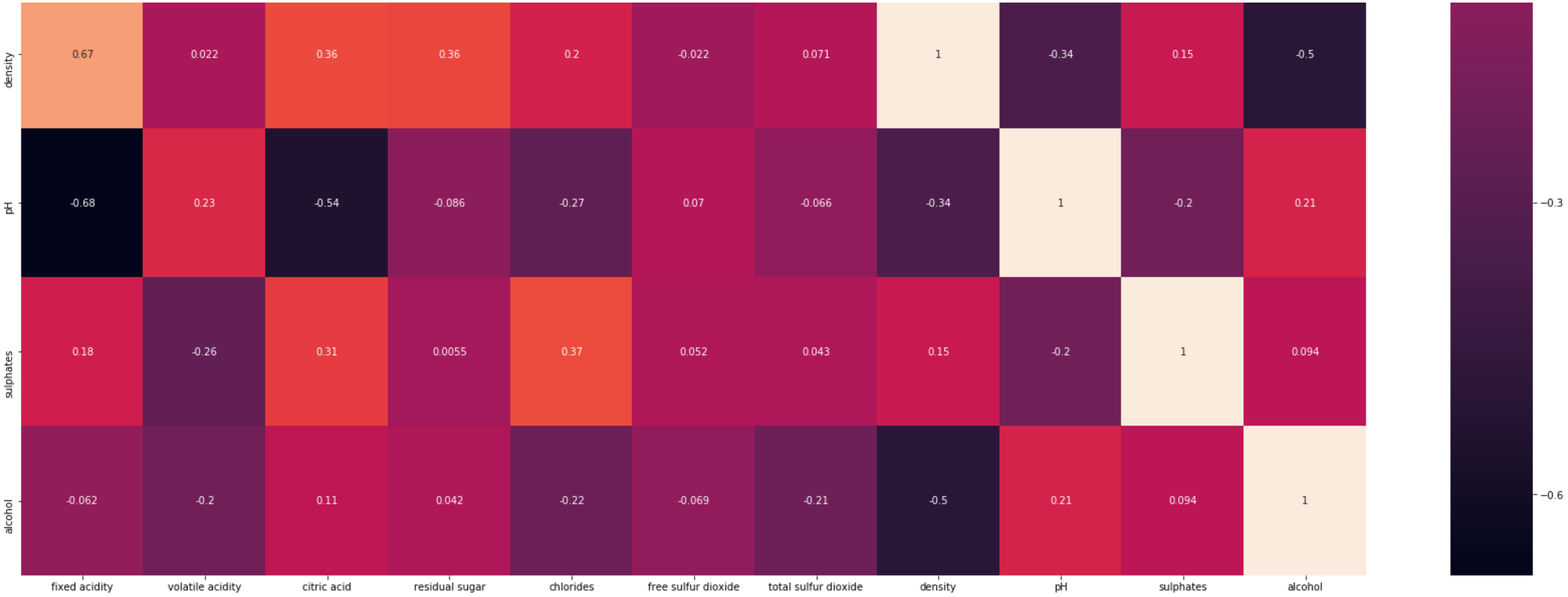
```
In [28]: #information about each numeric features
         print(X.describe())
```

```
       fixed acidity  volatile acidity  citric acid  residual sugar  \
count    1599.000000       1599.000000  1599.000000     1599.000000   
mean        8.319637          0.527821     0.270976        2.538806   
std         1.741096          0.179060     0.194801        1.409928   
min         4.600000          0.120000     0.000000        0.900000   
25%         7.100000          0.390000     0.090000        1.900000   
50%         7.900000          0.520000     0.260000        2.200000   
75%         9.200000          0.640000     0.420000        2.600000   
max        15.900000          1.580000     1.000000       15.500000   

         chlorides  free sulfur dioxide  total sulfur dioxide      density  \
count  1599.000000          1599.000000           1599.000000  1599.000000   
mean      0.087467            15.874922             46.467792     0.996747   
std       0.047065            10.460157             32.895324     0.001887   
min       0.012000             1.000000              6.000000     0.990070   
25%       0.070000             7.000000             22.000000     0.995600   
50%       0.079000            14.000000             38.000000     0.996750   
75%       0.090000            21.000000             62.000000     0.997835   
max       0.611000            72.000000            289.000000     1.003690   

                pH    sulphates      alcohol  
count  1599.000000  1599.000000  1599.000000  
mean      3.311113     0.658149    10.422983  
std       0.154386     0.169507     1.065668  
min       2.740000     0.330000     8.400000  
25%       3.210000     0.550000     9.500000  
50%       3.310000     0.620000    10.200000  
75%       3.400000     0.730000    11.100000  
max       4.010000     2.000000    14.900000  
```

In [29]:
```
plt.figure(figsize=[30,30])
sns.heatmap(X.corr(), annot = True)
```

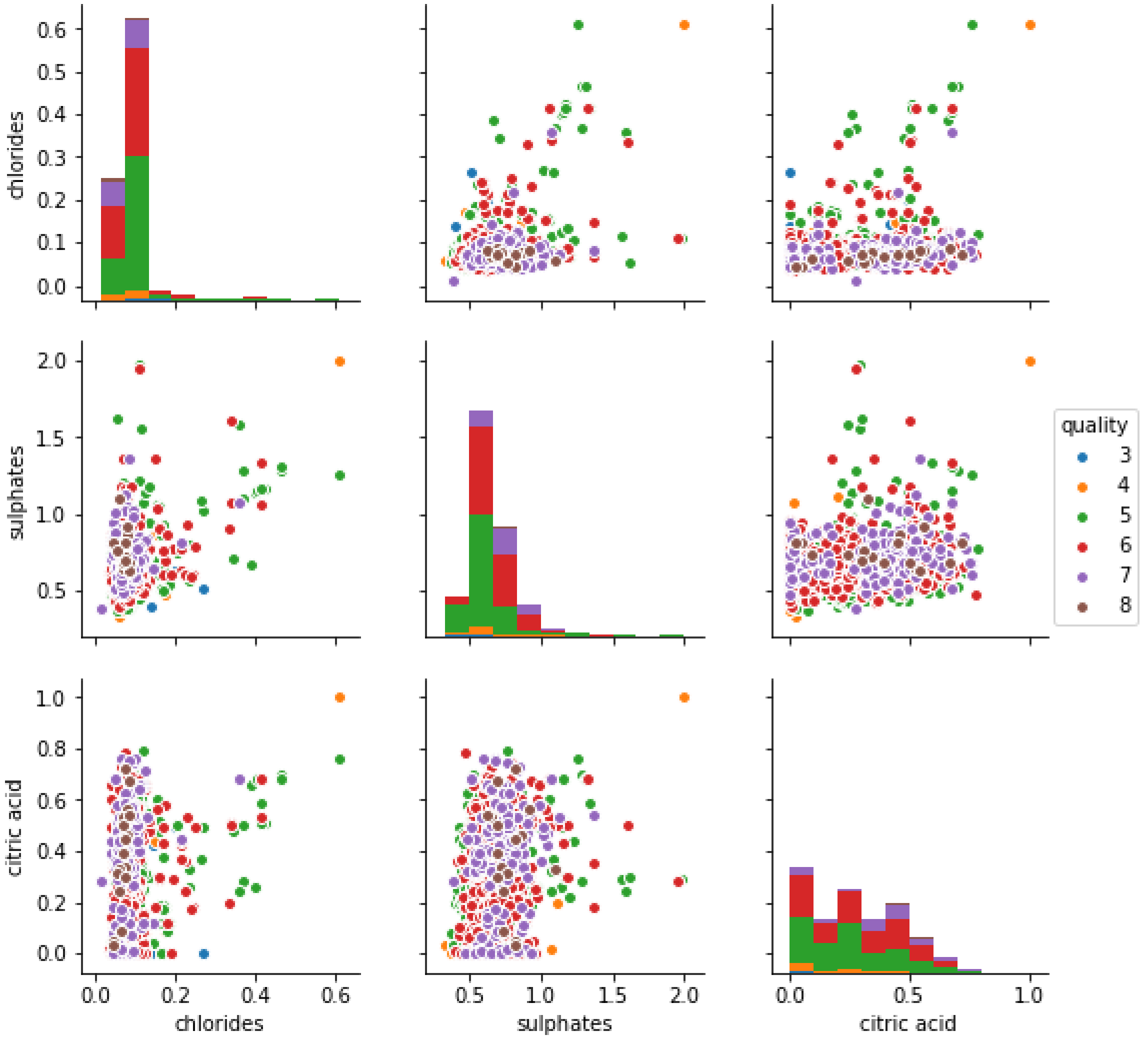
Out[29]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a21e186a0>
```




density
0.67
0.022
0.36
0.36
0.2
-0.022
0.071
1
-0.34
0.15
-0.5
pH
-0.68
0.23
-0.54
-0.086
-0.27
0.07
-0.066
-0.34
1
-0.2
0.21
sulphates
0.18
-0.26
0.31
0.0055
0.37
0.052
0.043
0.15
-0.2
1
0.094
alcohol
-0.062
-0.2
0.11
0.042
-0.22
-0.069
-0.21
-0.5
0.21
0.094
1
fixed acidity
volatile acidity
citric acid
residual sugar
chlorides
free sulfur dioxide
total sulfur dioxide
density
pH
sulphates
alcohol
−0.3
−0.6

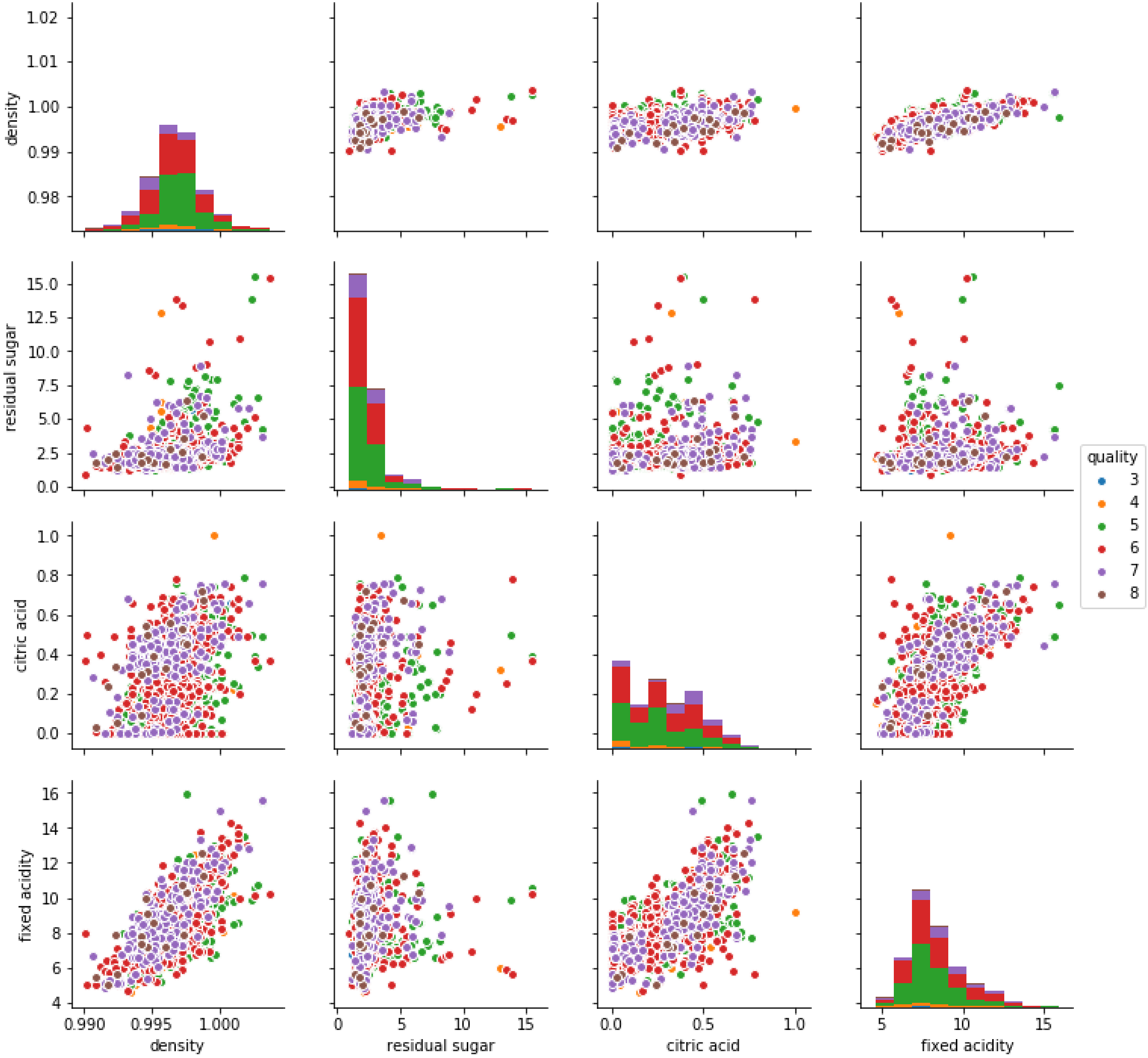
```
In [33]: sns.pairplot(mData,vars=["chlorides", "sulphates", "citric acid"],hue='quality')
```

```
Out[33]: <seaborn.axisgrid.PairGrid at 0x1a23d76358>
```



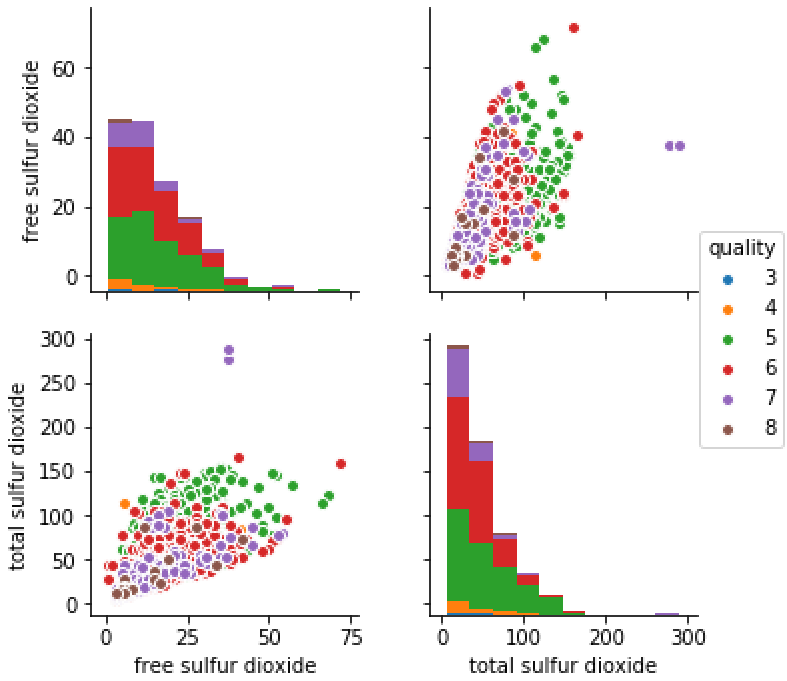
```
In [35]: sns.pairplot(mData,vars=["density", "residual sugar", "citric acid","fixed acidity"],hue='quality')
```

```
Out[35]: <seaborn.axisgrid.PairGrid at 0x1a24df9320>
```


density
1.02
1.01
1.00
0.99
0.98
residual sugar
15.0
12.5
10.0
7.5
5.0
2.5
0.0
citric acid
1.0
0.8
0.6
0.4
0.2
0.0
fixed acidity
16
14
12
10
8
6
4
0.990
0.995
1.000
density
0
5
10
15
residual sugar
0.0
0.5
1.0
citric acid
5
10
15
fixed acidity
quality
● 3
● 4
● 5
● 6
● 7
● 8

```
In [36]: sns.pairplot(mData,vars=["free sulfur dioxide", "total sulfur dioxide"],hue='quality')
```

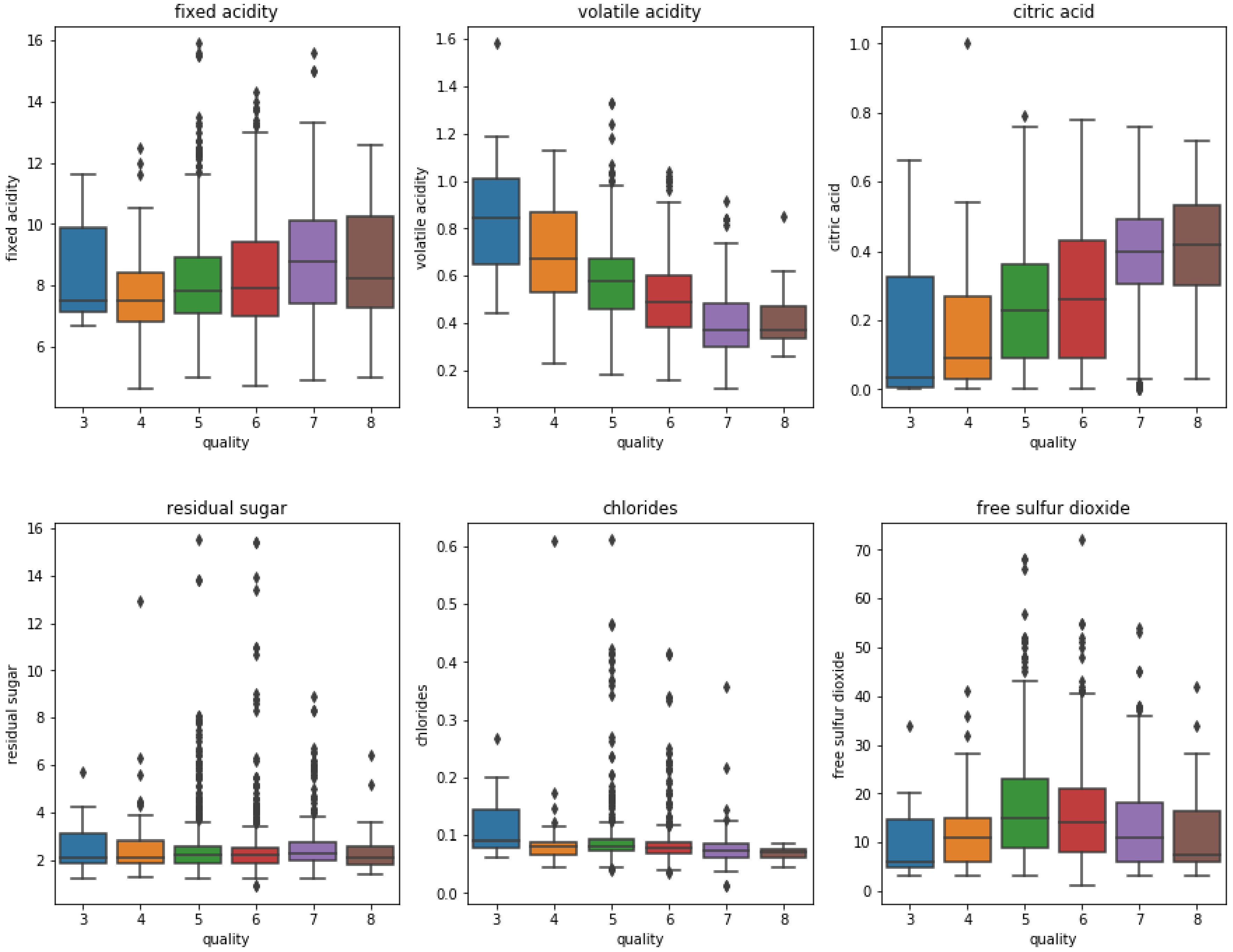
```
Out[36]: <seaborn.axisgrid.PairGrid at 0x1a25aa41d0>
```

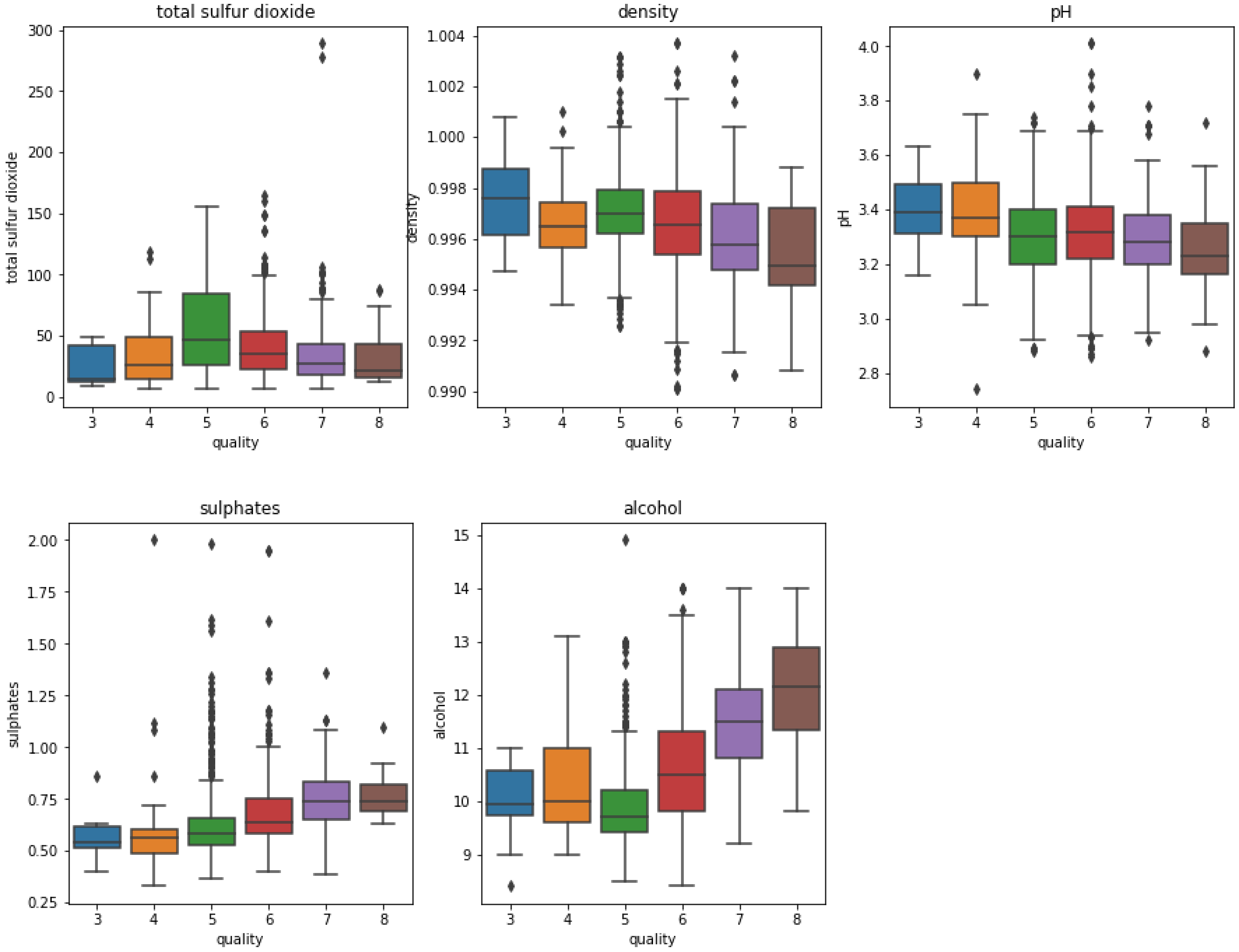


In [37]:

```
#histogram for numeric attributes
plt.figure(1,figsize=[15,5])
plt.subplot(1,3,1)
sns.boxplot(x = Y, y = X['fixed acidity'])
plt.title("fixed acidity")


plt.figure(1,figsize=[15,5])
plt.subplot(1,3,2)
sns.boxplot(x = Y, y = X['volatile acidity'])
plt.title("volatile acidity")

plt.figure(1,figsize=[15,5])
plt.subplot(1,3,3)
sns.boxplot(x = Y, y = X['citric acid'])
plt.title("citric acid")

plt.figure(2,figsize=[15,5])
plt.subplot(1,3,1)
sns.boxplot(x = Y, y = X['residual sugar'])
plt.title("residual sugar")


plt.figure(2,figsize=[15,5])
plt.subplot(1,3,2)
sns.boxplot(x = Y, y = X['chlorides'])
plt.title("chlorides")


plt.figure(2,figsize=[15,5])
plt.subplot(1,3,3)
sns.boxplot(x = Y, y = X['free sulfur dioxide'])
plt.title("free sulfur dioxide")

plt.figure(3,figsize=[15,5])
plt.subplot(1,3,1)
sns.boxplot(x = Y, y = X['total sulfur dioxide'])
plt.title("total sulfur dioxide")


plt.figure(3,figsize=[15,5])
plt.subplot(1,3,2)
sns.boxplot(x = Y, y = X['density'])
plt.title("density")


plt.figure(3,figsize=[15,5])
plt.subplot(1,3,3)
sns.boxplot(x = Y, y = X['pH'])
plt.title("pH")

plt.figure(4,figsize=[15,5])
plt.subplot(1,3,1)
sns.boxplot(x = Y, y = X['sulphates'])
plt.title("sulphates")

plt.figure(4,figsize=[15,5])
plt.subplot(1,3,2)
sns.boxplot(x = Y, y = X['alcohol'])
plt.title("alcohol")
```

```
Out[37]: Text(0.5,1,'alcohol')
```


fixed acidity
fixed acidity
16
14
12
10
8
6
quality
volatile acidity
volatile acidity
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
quality
citric acid
citric acid
1.0
0.8
0.6
0.4
0.2
0.0
quality
residual sugar
residual sugar
16
14
12
10
8
6
4
2
quality
chlorides
chlorides
0.6
0.5
0.4
0.3
0.2
0.1
0.0
quality
free sulfur dioxide
free sulfur dioxide
70
60
50
40
30
20
10
0
quality
3
4
5
6
7
8


total sulfur dioxide
total sulfur dioxide
0
50
100
150
200
250
300
3
4
5
6
7
8
quality
density
density
0.990
0.992
0.994
0.996
0.998
1.000
1.002
1.004
3
4
5
6
7
8
quality
pH
pH
2.8
3.0
3.2
3.4
3.6
3.8
4.0
3
4
5
6
7
8
quality
sulphates
sulphates
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
3
4
5
6
7
8
quality
alcohol
alcohol
9
10
11
12
13
14
15
3
4
5
6
7
8
quality

In [ ]: