

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

```
df=pd.read_csv("/content/WineQT.csv")
df
```



	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality	Id
0	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	0.56	9.4	5	0
1	7.8	0.880	0.00	2.6	0.098	25.0	67.0	0.99680	3.20	0.68	9.8	5	1
2	7.8	0.760	0.04	2.3	0.092	15.0	54.0	0.99700	3.26	0.65	9.8	5	2
3	11.2	0.280	0.56	1.9	0.075	17.0	60.0	0.99800	3.16	0.58	9.8	6	3
4	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	0.56	9.4	5	4
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1138	6.3	0.510	0.13	2.3	0.076	29.0	40.0	0.99574	3.42	0.75	11.0	6	1592
1139	6.8	0.620	0.08	1.9	0.068	28.0	38.0	0.99651	3.42	0.82	9.5	6	1593
1140	6.2	0.600	0.08	2.0	0.090	32.0	44.0	0.99490	3.45	0.58	10.5	5	1594
1141	5.9	0.550	0.10	2.2	0.062	39.0	51.0	0.99512	3.52	0.76	11.2	6	1595
1142	5.9	0.645	0.12	2.0	0.075	32.0	44.0	0.99547	3.57	0.71	10.2	5	1597

1143 rows × 13 columns

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1143 entries, 0 to 1142
Data columns (total 13 columns):
#   Column              Non-Null Count  Dtype
---  -
#   Column              Non-Null Count  Dtype
```

```

0    fixed acidity    1143 non-null    float64
1    volatile acidity 1143 non-null    float64
2    citric acid      1143 non-null    float64
3    residual sugar   1143 non-null    float64
4    chlorides        1143 non-null    float64
5    free sulfur dioxide 1143 non-null    float64
6    total sulfur dioxide 1143 non-null    float64
7    density          1143 non-null    float64
8    pH              1143 non-null    float64
9    sulphates        1143 non-null    float64
10   alcohol          1143 non-null    float64
11   quality          1143 non-null    int64
12   Id              1143 non-null    int64

```

```

dtypes: float64(11), int64(2)
memory usage: 116.2 KB

```

```
df.head()
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality	Id
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5	0
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8	5	1
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9.8	5	2
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8	6	3
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5	4

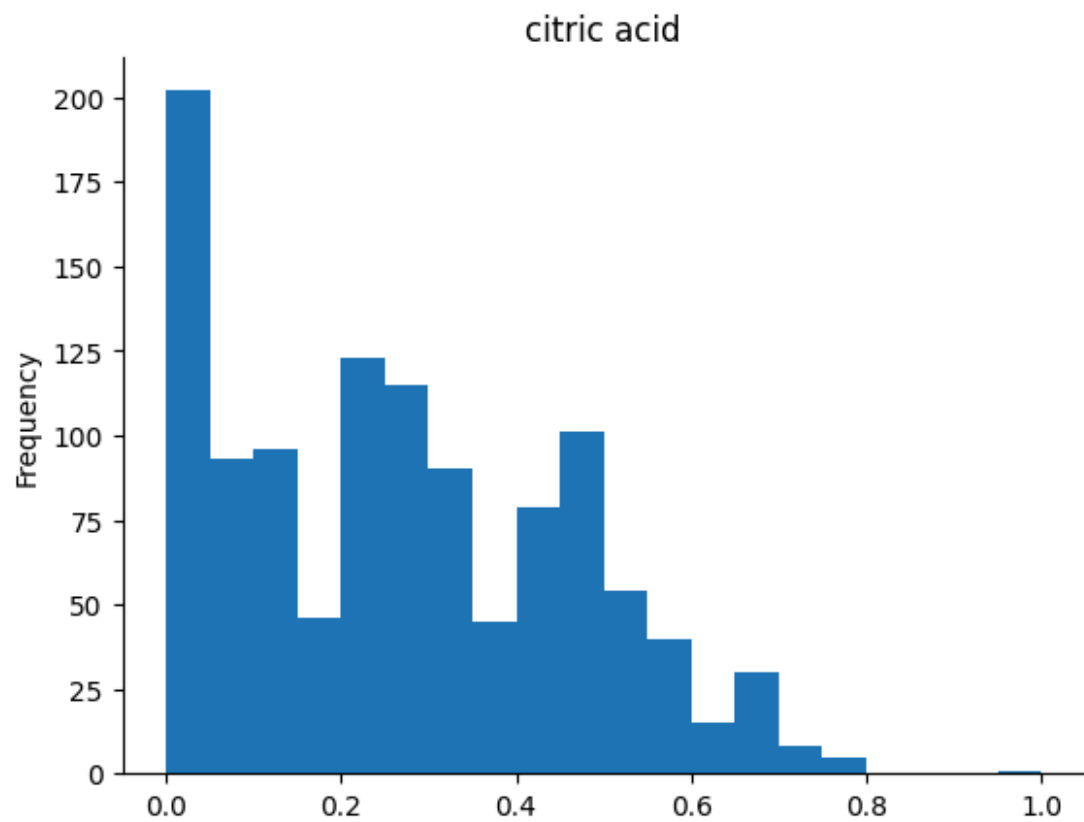
## ▼ citric acid

```
# @title citric acid
```

```

from matplotlib import pyplot as plt
df['citric acid'].plot(kind='hist', bins=20, title='citric acid')
plt.gca().spines[['top', 'right',]].set_visible(False)

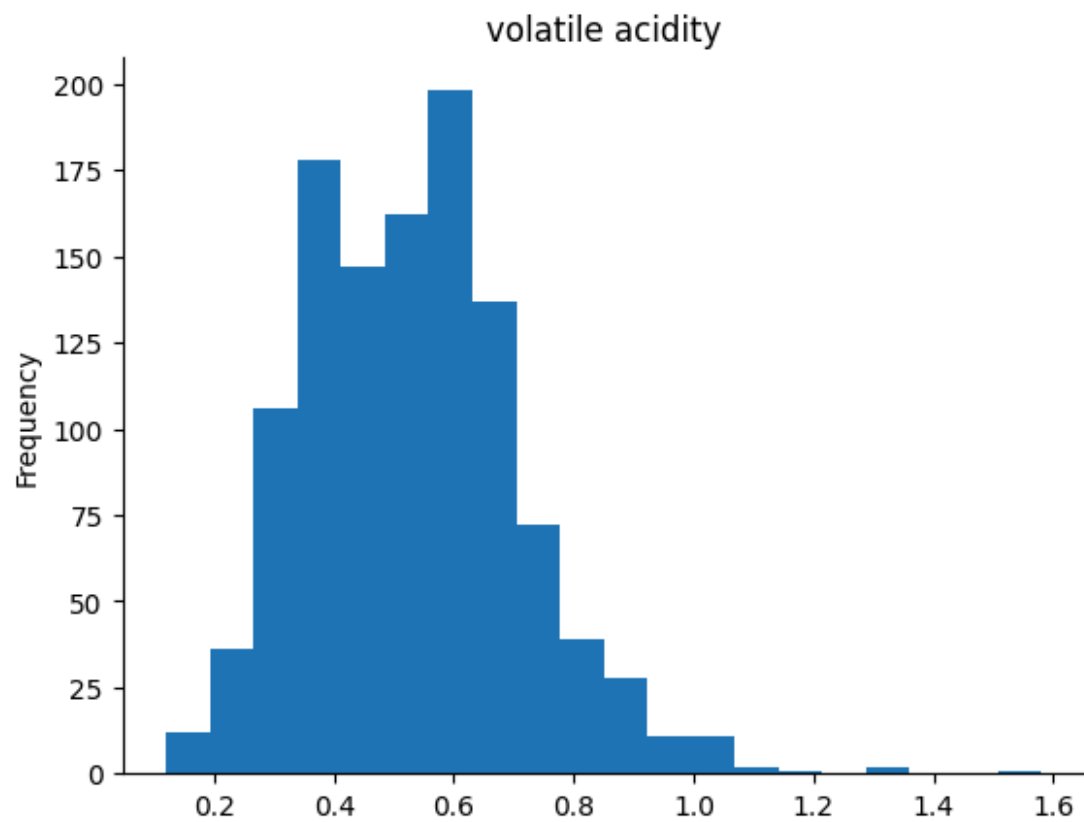
```



## ▼ volatile acidity

```
# @title volatile acidity
```

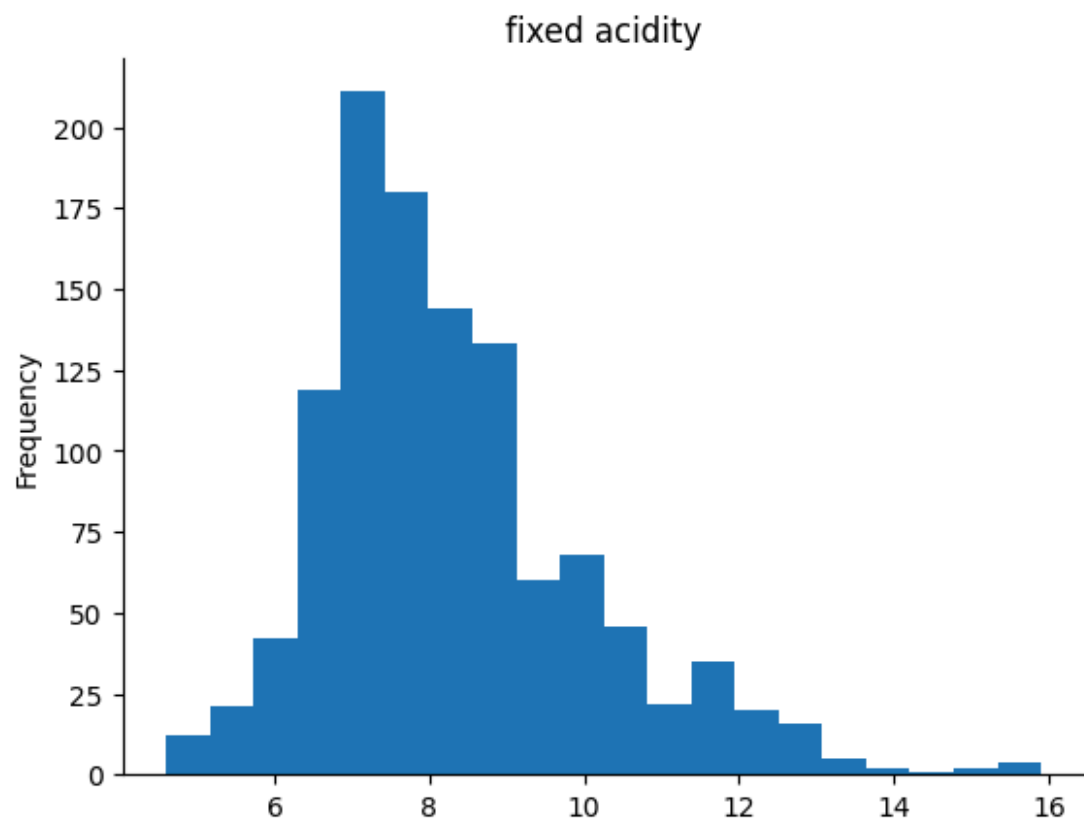
```
from matplotlib import pyplot as plt
df['volatile acidity'].plot(kind='hist', bins=20, title='volatile acidity')
plt.gca().spines[['top', 'right']].set_visible(False)
```



## ▼ fixed acidity

```
# @title fixed acidity
```

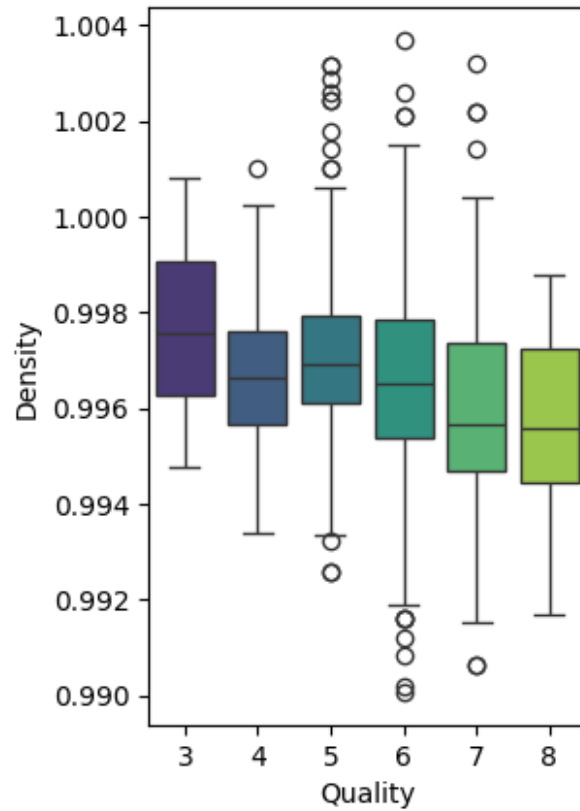
```
from matplotlib import pyplot as plt
df['fixed acidity'].plot(kind='hist', bins=20, title='fixed acidity')
plt.gca().spines[['top', 'right']].set_visible(False)
```



```
plt.subplot(1, 2, 1)
sns.boxplot(x='quality', y='density', data=df, palette='viridis')
plt.title('Density across Wine Quality Levels')
plt.xlabel('Quality')
plt.ylabel('Density')
```

Passing ``palette`` without assigning ``hue`` is deprecated and will be removed in v0.14.0. Assign the ``x`` variable to ``hue`` and set ``legend=F``

### Density across Wine Quality Levels



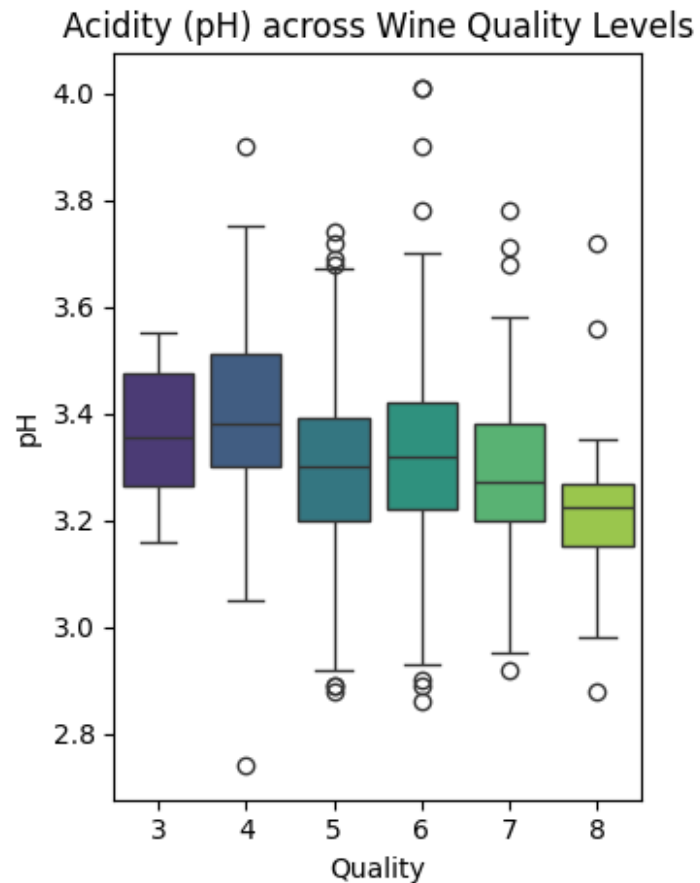
```
# Acidity distribution across different quality levels
plt.subplot(1, 2, 2)
sns.boxplot(x='quality', y='pH', data=df, palette='viridis')
plt.title('Acidity (pH) across Wine Quality Levels')
plt.xlabel('Quality')
plt.ylabel('pH')

plt.tight_layout()
plt.show()
```

<ipython-input-43-4ac52eaba85b>:3: FutureWarning:

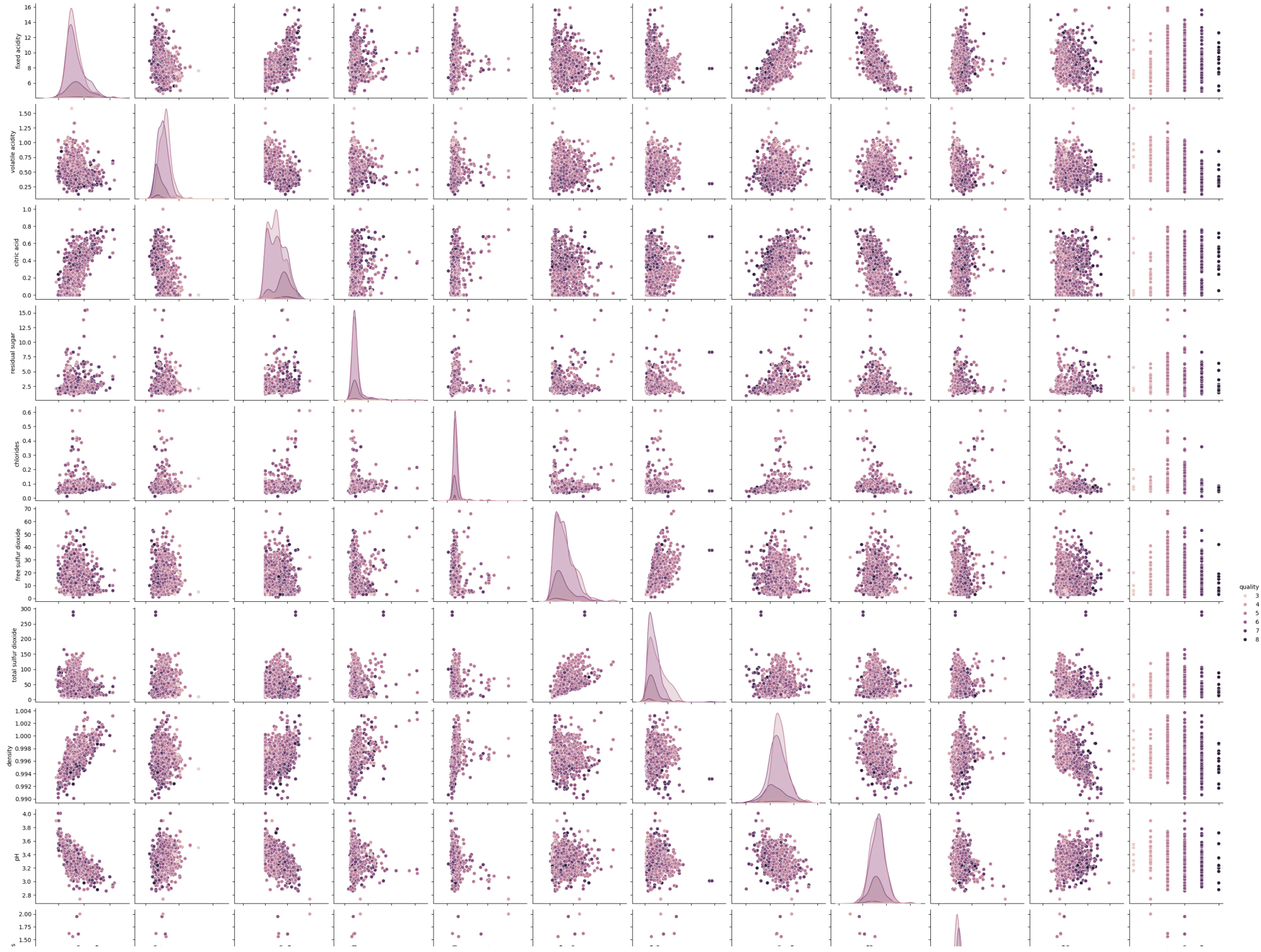
Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False`.

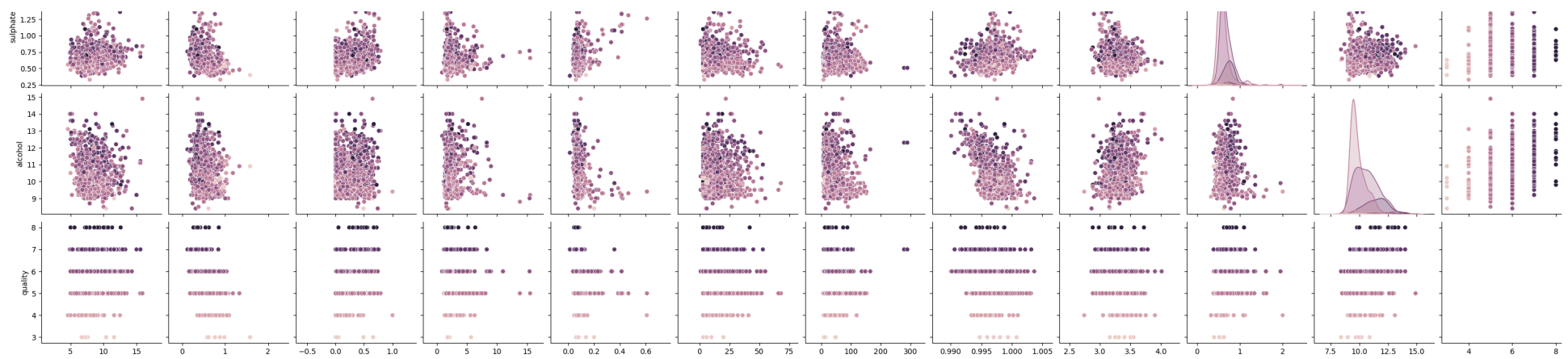
```
sns.boxplot(x='quality', y='pH', data=df, palette='viridis')
```



```
sns.pairplot(df, vars=['fixed acidity', 'volatile acidity', 'citric acid',  
                      'residual sugar', 'chlorides', 'free sulfur dioxide',  
                      'total sulfur dioxide', 'density', 'pH', 'sulphates', 'alcohol', 'quality'],  
             hue='quality', diag_kind='kde')  
  
plt.show()
```

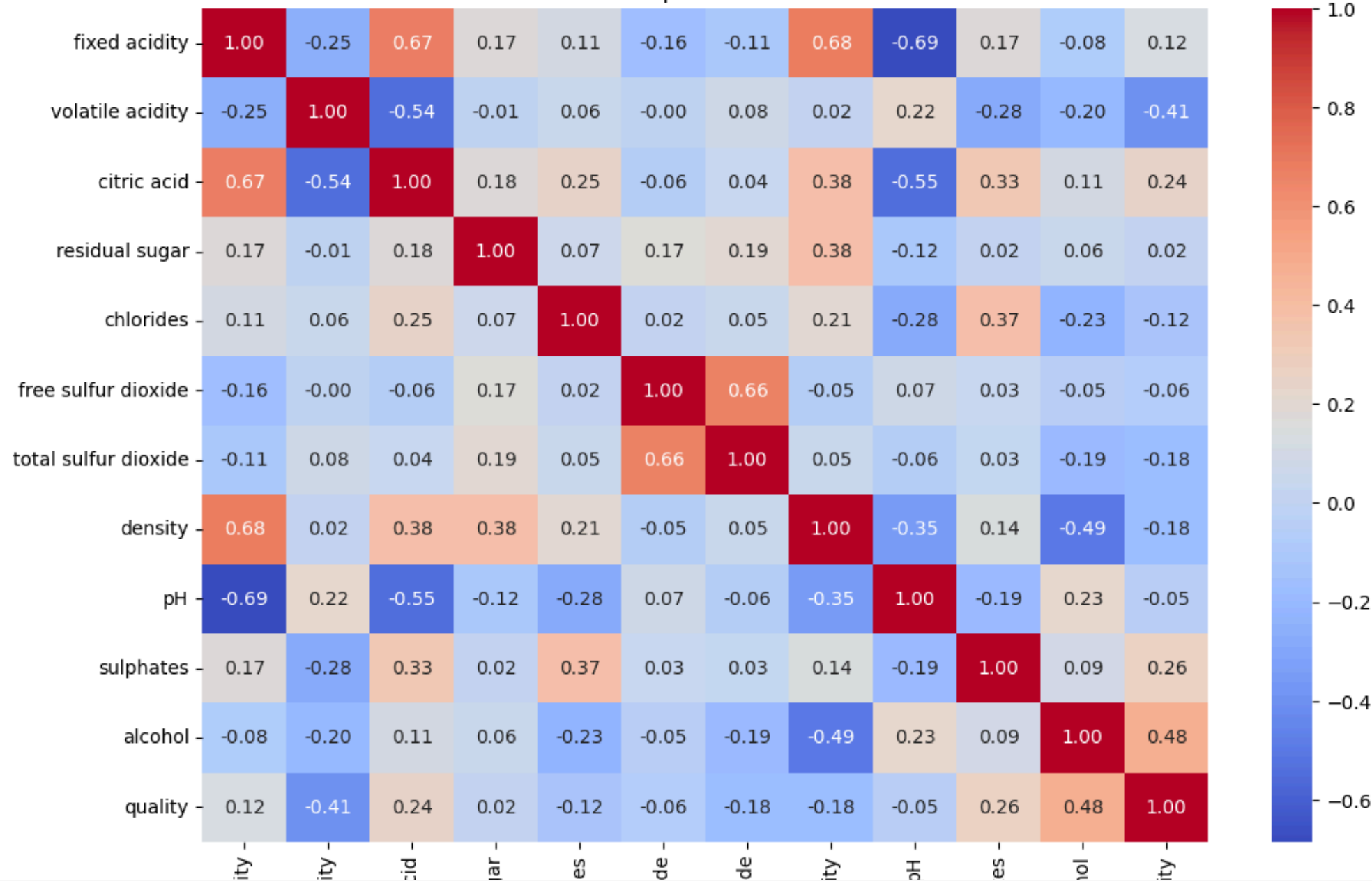






```
# Correlation matrix heatmap
plt.figure(figsize=(12, 8))
sns.heatmap(df.corr(), annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Heatmap of Wine Dataset Features')
plt.show()
```

Correlation Heatmap of Wine Dataset Features

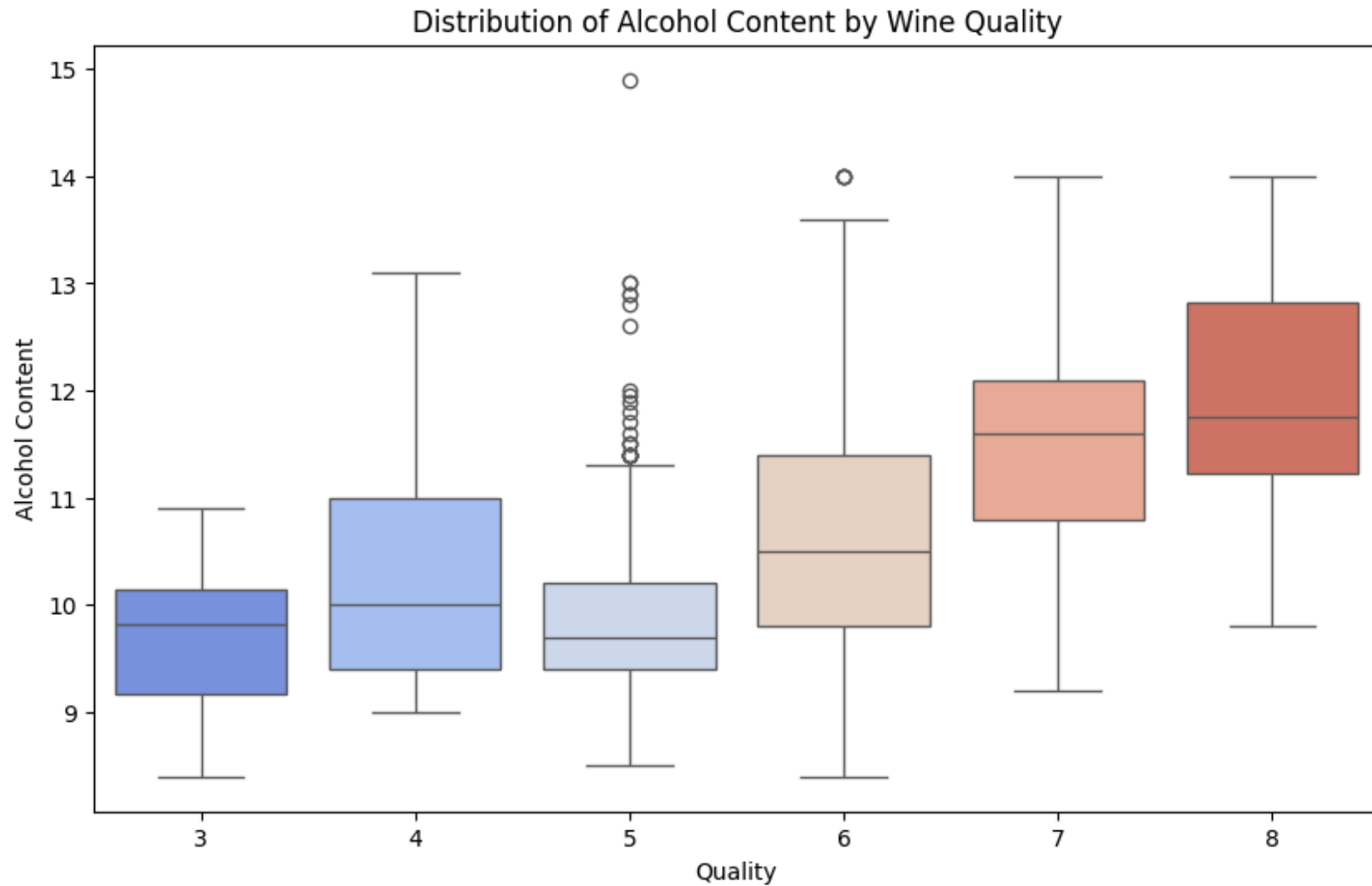


```
# Boxplot of wine quality distribution
plt.figure(figsize=(10, 6))
sns.boxplot(x='quality', y='alcohol', data=df, palette='coolwarm')
plt.title('Distribution of Alcohol Content by Wine Quality')
plt.xlabel('Quality')
plt.ylabel('Alcohol Content')
plt.show()
```

```
<ipython-input-49-92fea6fb7240>:3: FutureWarning:
```

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=f

```
sns.boxplot(x='quality', y='alcohol', data=df, palette='coolwarm')
```



```
df.tail()
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality	Id
<b>1138</b>	6.3	0.510	0.13	2.3	0.076	29.0	40.0	0.99574	3.42	0.75	11.0	6	1592
<b>1139</b>	6.8	0.620	0.08	1.9	0.068	28.0	38.0	0.99651	3.42	0.82	9.5	6	1593
<b>1140</b>	6.2	0.600	0.08	2.0	0.090	32.0	44.0	0.99490	3.45	0.58	10.5	5	1594
<b>1141</b>	5.9	0.550	0.10	2.2	0.062	39.0	51.0	0.99512	3.52	0.76	11.2	6	1595
<b>1142</b>	5.9	0.645	0.12	2.0	0.075	32.0	44.0	0.99547	3.57	0.71	10.2	5	1597

```
row,columns=df.shape
print("The number of rows are",row)
print("The number of columns are",columns)
```

```
The number of rows are 1143
The number of columns are 13
```

```
df.size
```

```
14859
```

```
df.columns
```

```
Index(['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar',
      'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'density',
      'pH', 'sulphates', 'alcohol', 'quality', 'Id'],
      dtype='object')
```

```
df.dtypes
```

```
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
Id                     int64
dtype: object
```

```
df.isna().sum()
```

```
fixed acidity          0
volatile acidity       0
citric acid            0
residual sugar         0
chlorides              0
free sulfur dioxide    0
total sulfur dioxide    0
density               0
pH                    0
sulphates             0
alcohol               0
quality               0
Id                    0
dtype: int64
```

```
df.drop('Id', inplace=True, axis=1)
```

```
from sklearn.model_selection import train_test_split
```

```
X = df.drop('quality', axis = 1)
```