model-summary-and-eda-1

February 6, 2024

LOADING DATA FROM GITHUB INTO SINGLESTORE NOTEBOOK

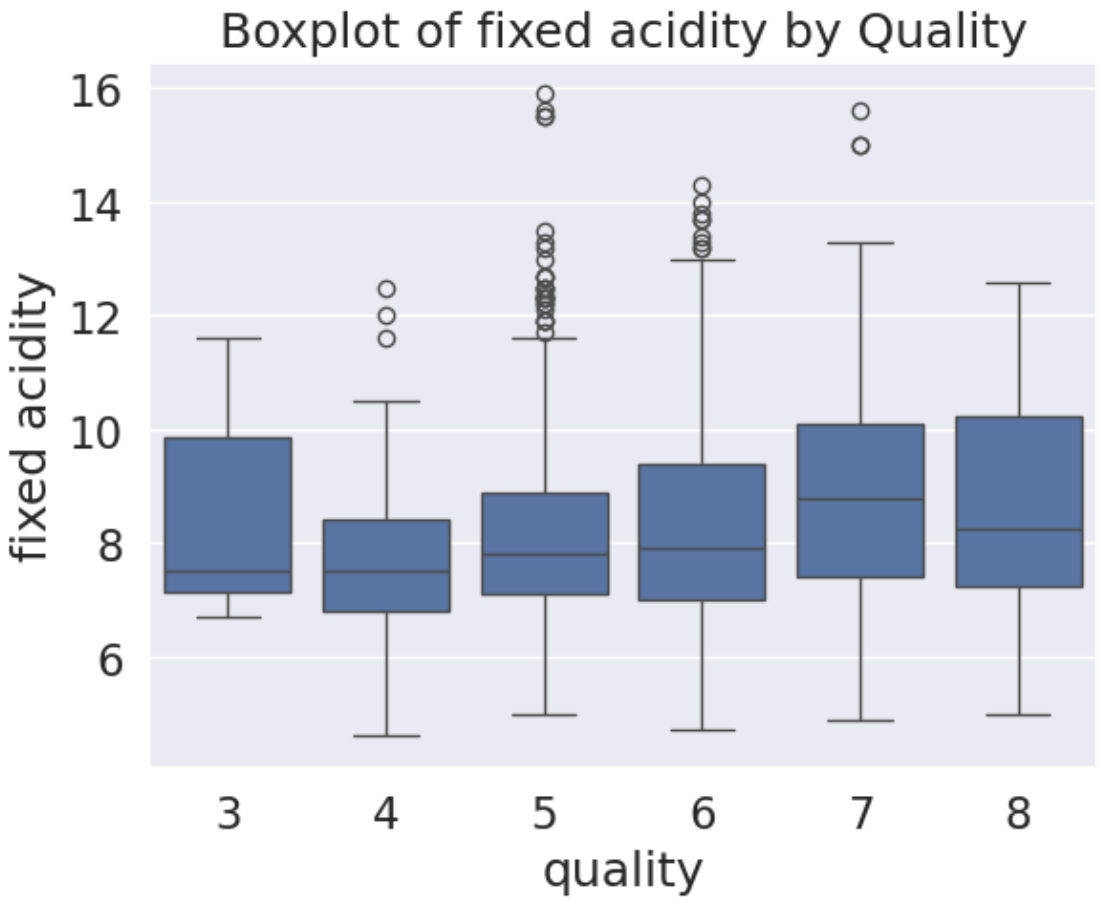
[4272]:
```
import pandas as pd

github_url = 'https://raw.githubusercontent.com/aniruddhachoudhury/
  ↪Red-Wine-Quality/master/winequality-red.csv'
wine_data = pd.read_csv(github_url)
print('Wine Quality Dataset:')
print(wine_data.head())
```

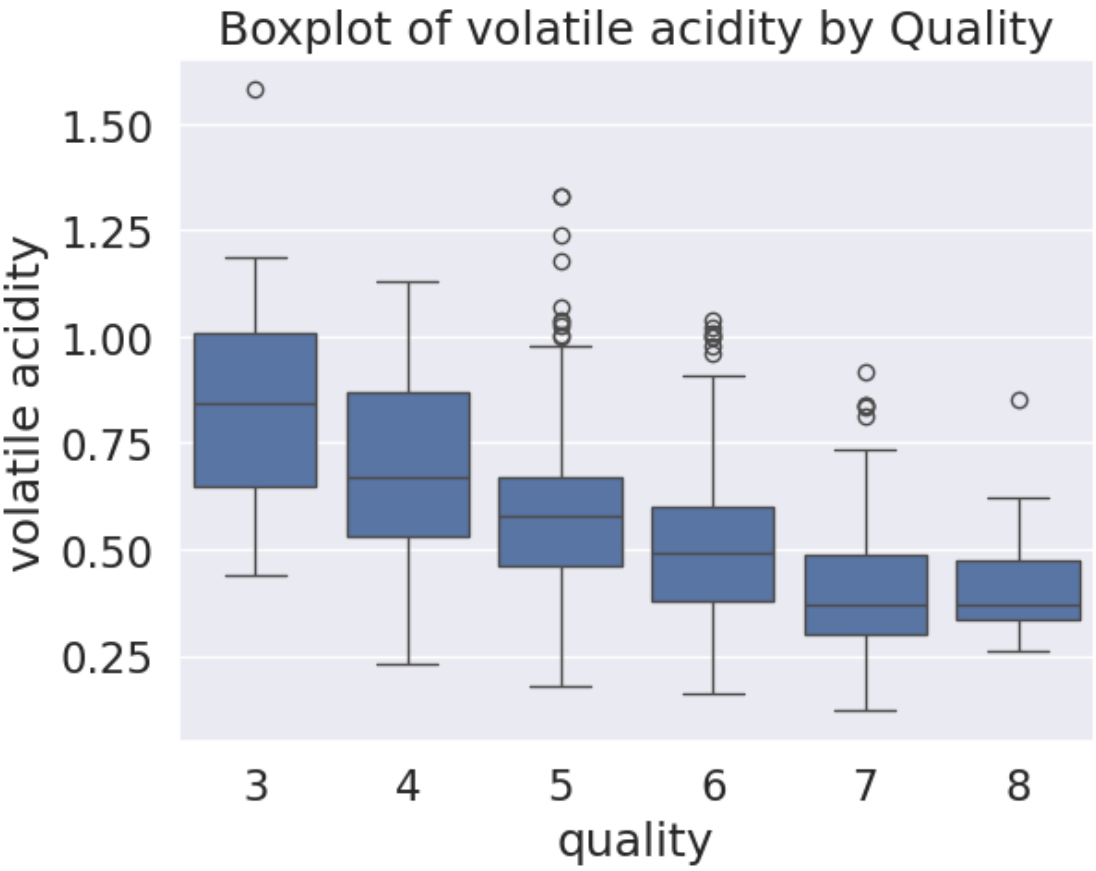
```
Wine Quality Dataset:
   fixed acidity  volatile acidity  citric acid  residual sugar  chlorides  \
0            7.4              0.70         0.00             1.9      0.076
1            7.8              0.88         0.00             2.6      0.098
2            7.8              0.76         0.04             2.3      0.092
3           11.2              0.28         0.56             1.9      0.075
4            7.4              0.70         0.00             1.9      0.076

   free sulfur dioxide  total sulfur dioxide  density    pH  sulphates  \
0                 11.0                  34.0   0.9978  3.51       0.56
1                 25.0                  67.0   0.9968  3.20       0.68
2                 15.0                  54.0   0.9970  3.26       0.65
3                 17.0                  60.0   0.9980  3.16       0.58
4                 11.0                  34.0   0.9978  3.51       0.56

   alcohol  quality
0      9.4        5
1      9.8        5
2      9.8        5
3      9.8        6
4      9.4        5
```

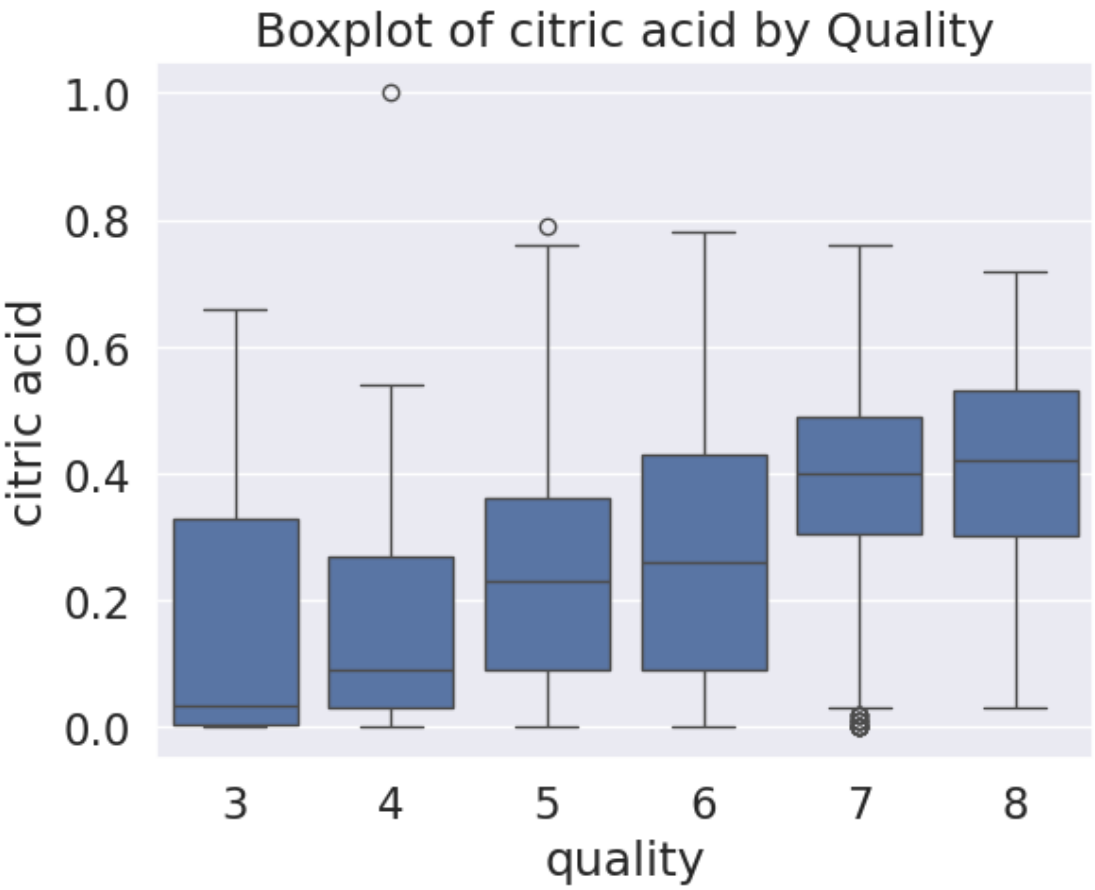
[4273]:
```
import seaborn as sns
import matplotlib.pyplot as plt

sns.histplot(wine_data['quality'], bins=range(3, 10), kde=False)
plt.xlabel('Quality')
plt.ylabel('Frequency')
```

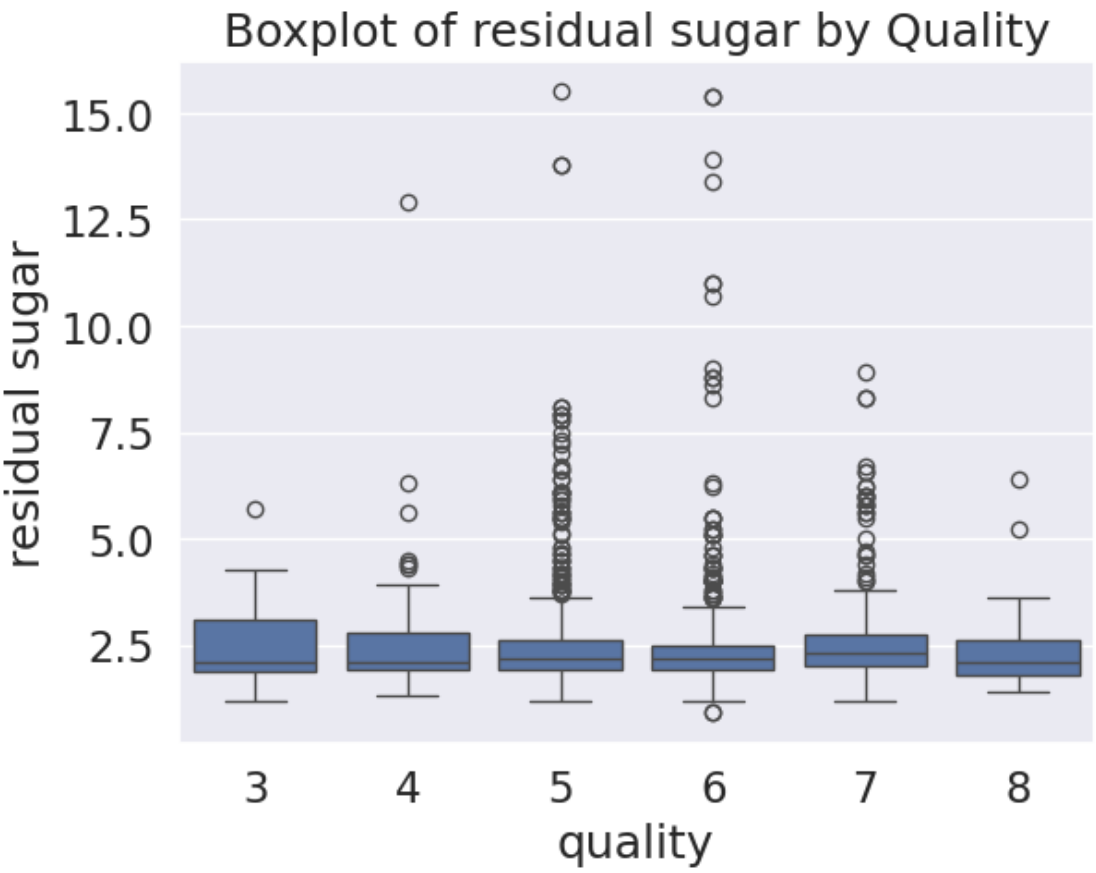
```
plt.title('Histogram of Quality')
plt.show()
```

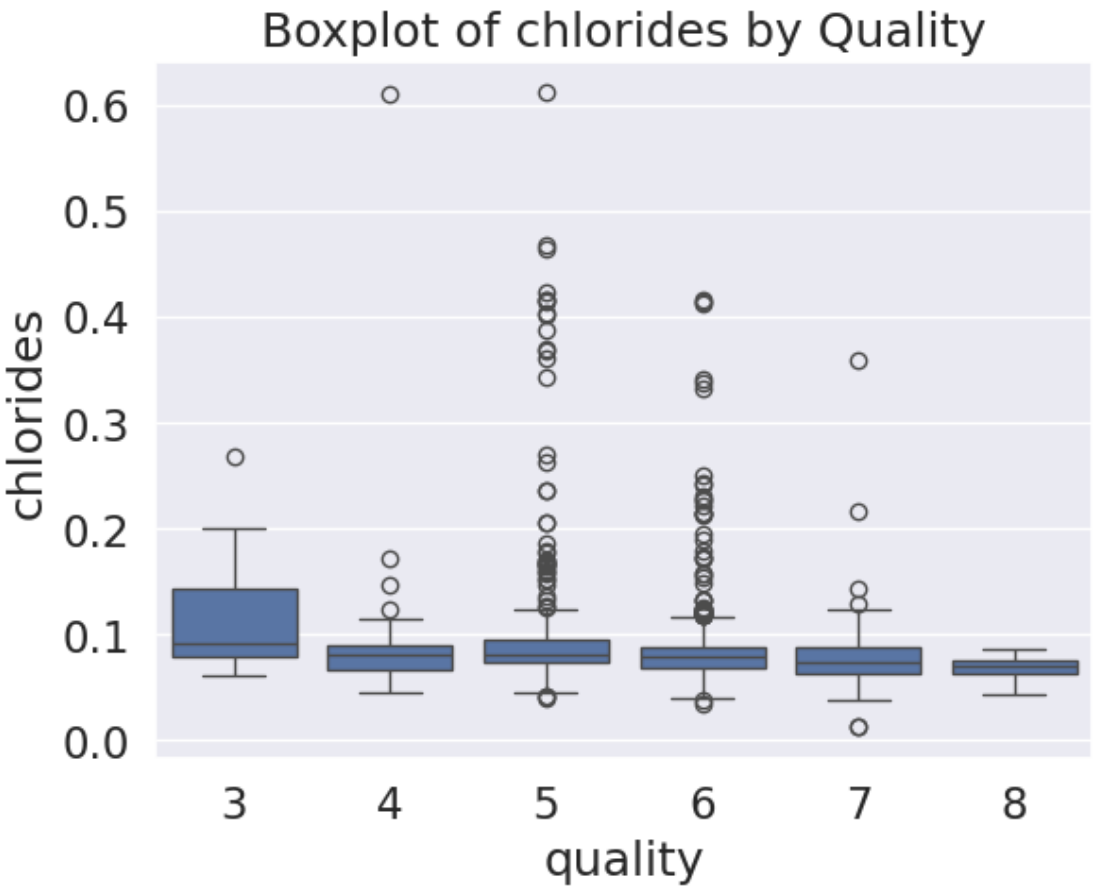
[4274]:
```
import seaborn as sns
import matplotlib.pyplot as plt

predictors = ['fixed acidity', 'volatile acidity', 'citric acid', 'residual␣
 ↪sugar', 'chlorides', 'free sulfur dioxide', 'total sulfur dioxide',␣
 ↪'density', 'pH', 'sulphates', 'alcohol']

for predictor in predictors:
    sns.boxplot(x='quality', y=predictor, data=wine_data)
    plt.title(f'Boxplot of {predictor} by Quality')
    plt.show()
```

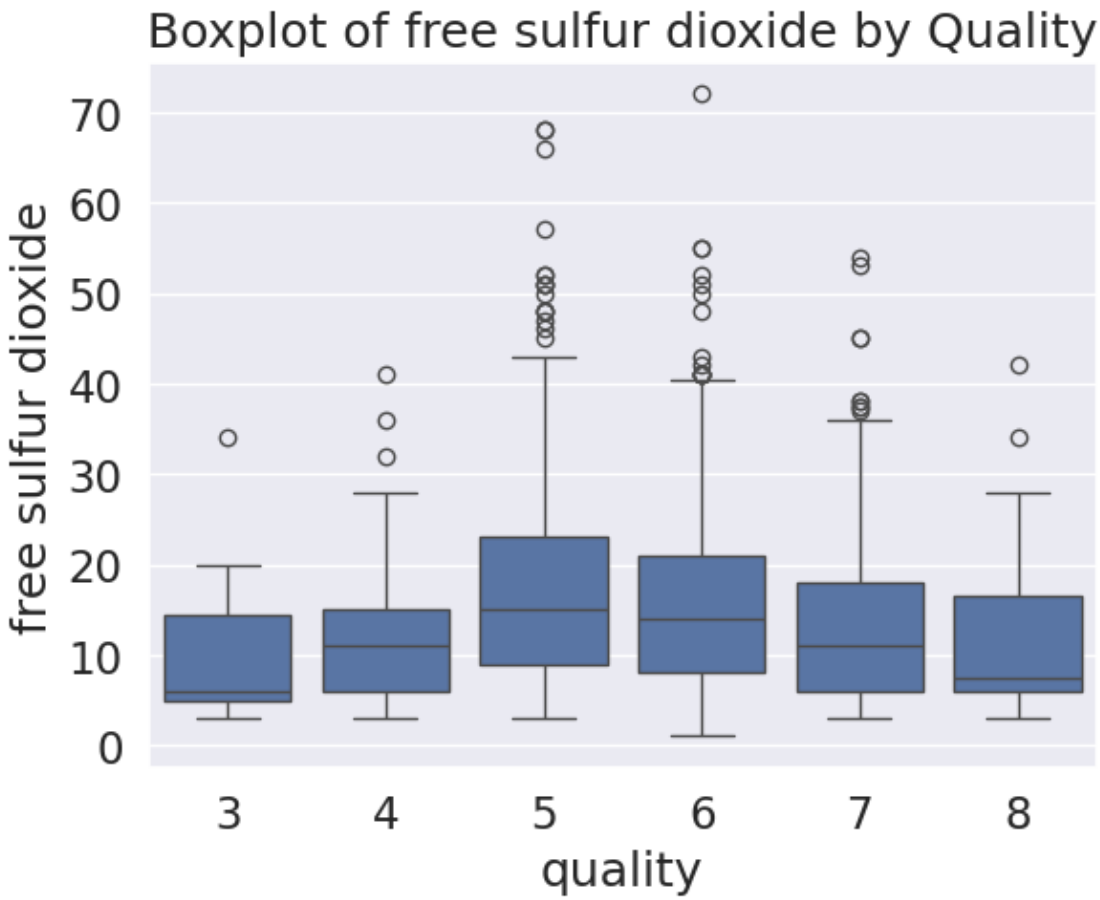

Boxplot of fixed acidity by Quality
fixed acidity
16
14
12
10
8
6
3
4
5
6
7
8
quality

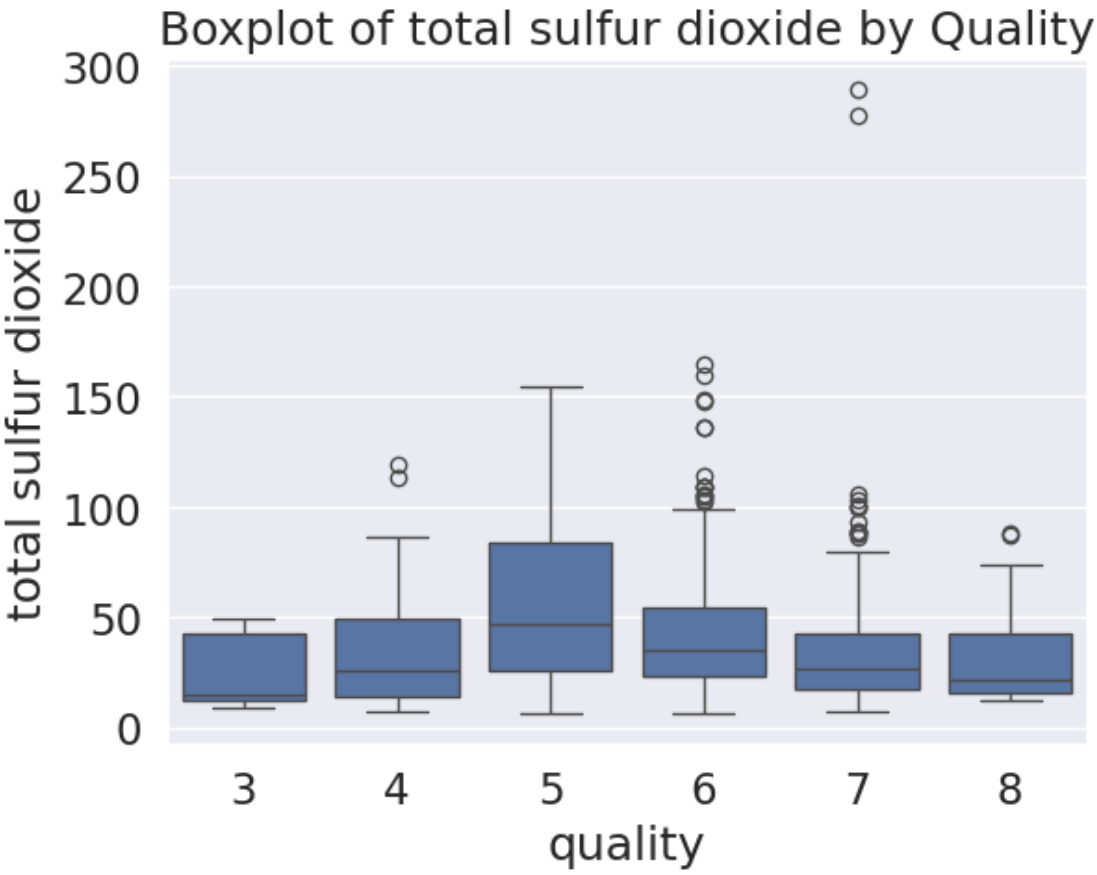

Boxplot of volatile acidity by Quality
volatile acidity
1.50
1.25
1.00
0.75
0.50
0.25
3
4
5
6
7
8
quality

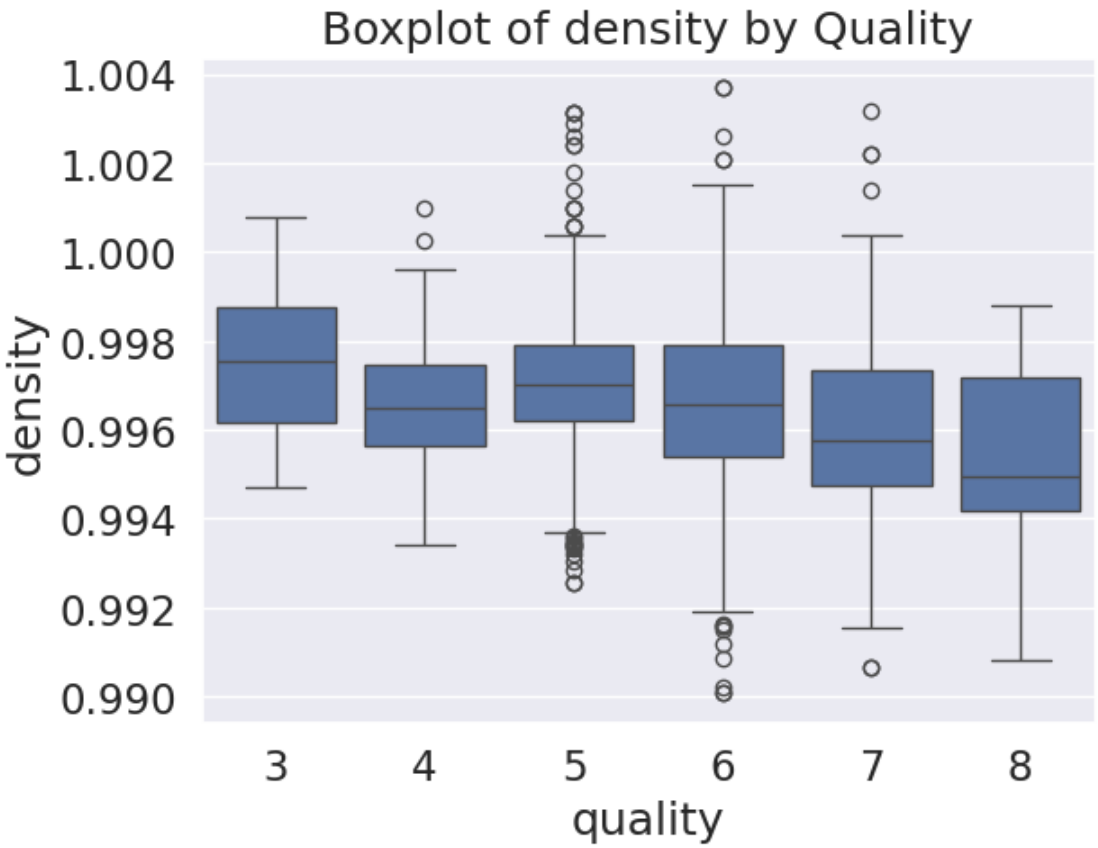

Boxplot of citric acid by Quality
citric acid
1.0
0.8
0.6
0.4
0.2
0.0
3
4
5
6
7
8
quality

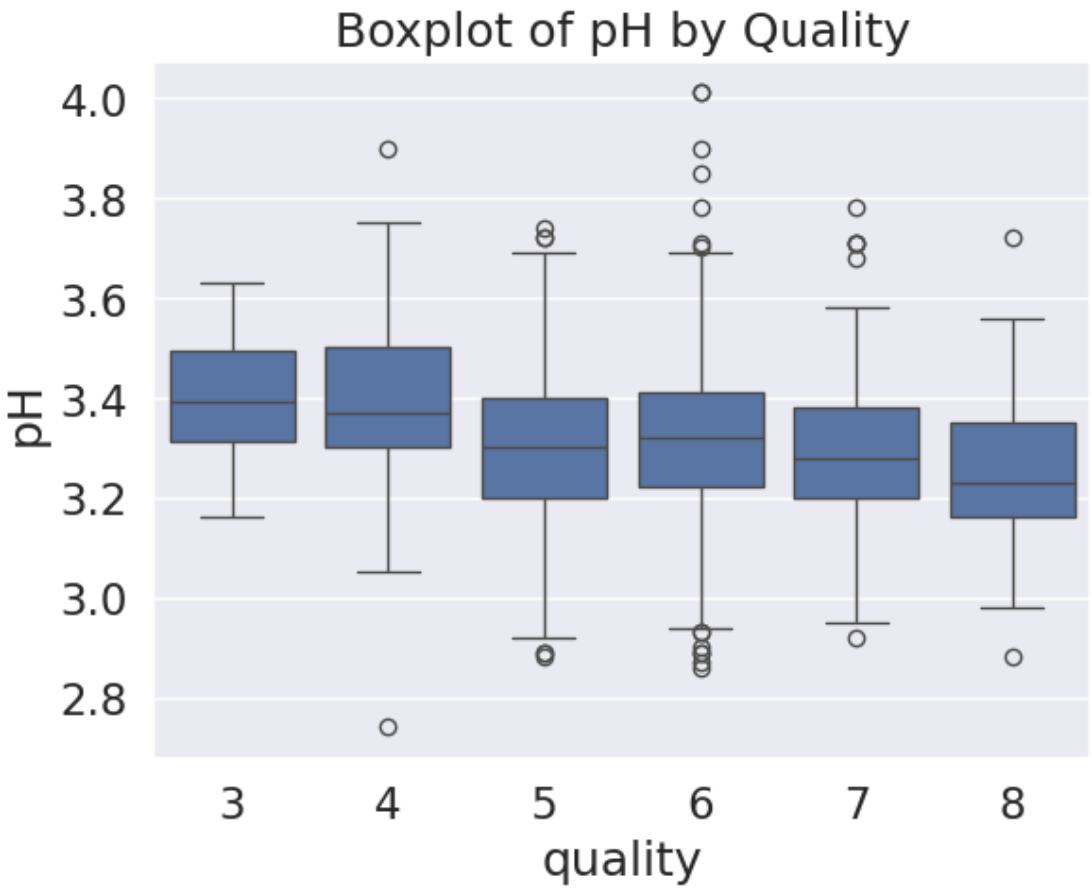

Boxplot of residual sugar by Quality
residual sugar
15.0
12.5
10.0
7.5
5.0
2.5
3
4
5
6
7
8
quality

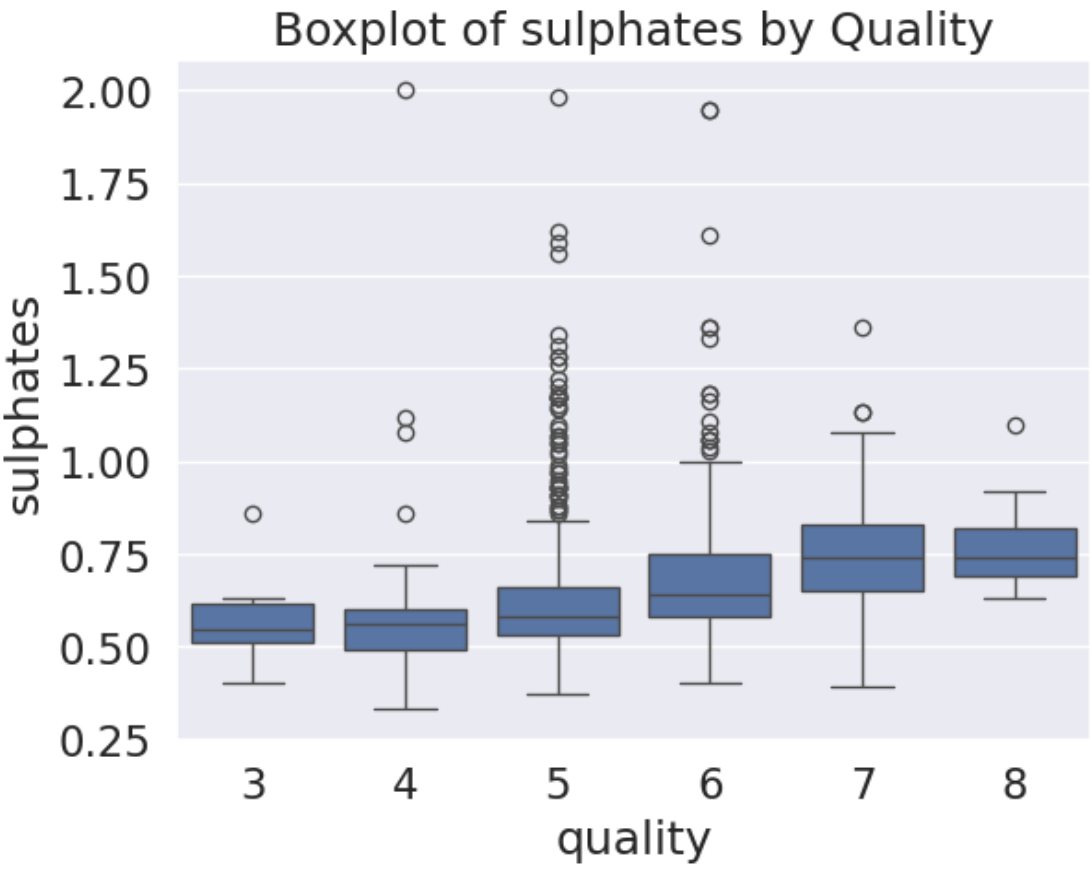

Boxplot of chlorides by Quality
chlorides
0.0
0.1
0.2
0.3
0.4
0.5
0.6
3
4
5
6
7
8
quality

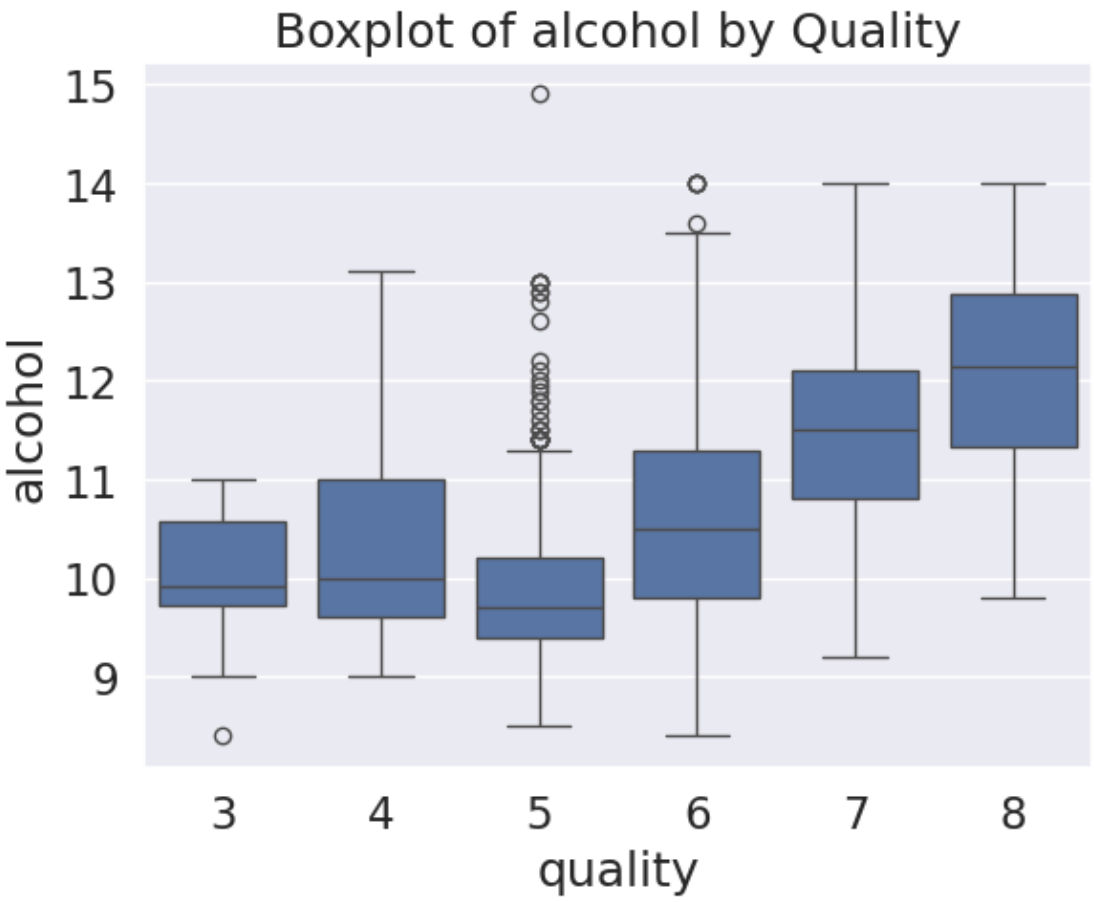

Boxplot of free sulfur dioxide by Quality
free sulfur dioxide
0
10
20
30
40
50
60
70
3
4
5
6
7
8
quality


Boxplot of total sulfur dioxide by Quality
total sulfur dioxide
300
250
200
150
100
50
0
3
4
5
6
7
8
quality

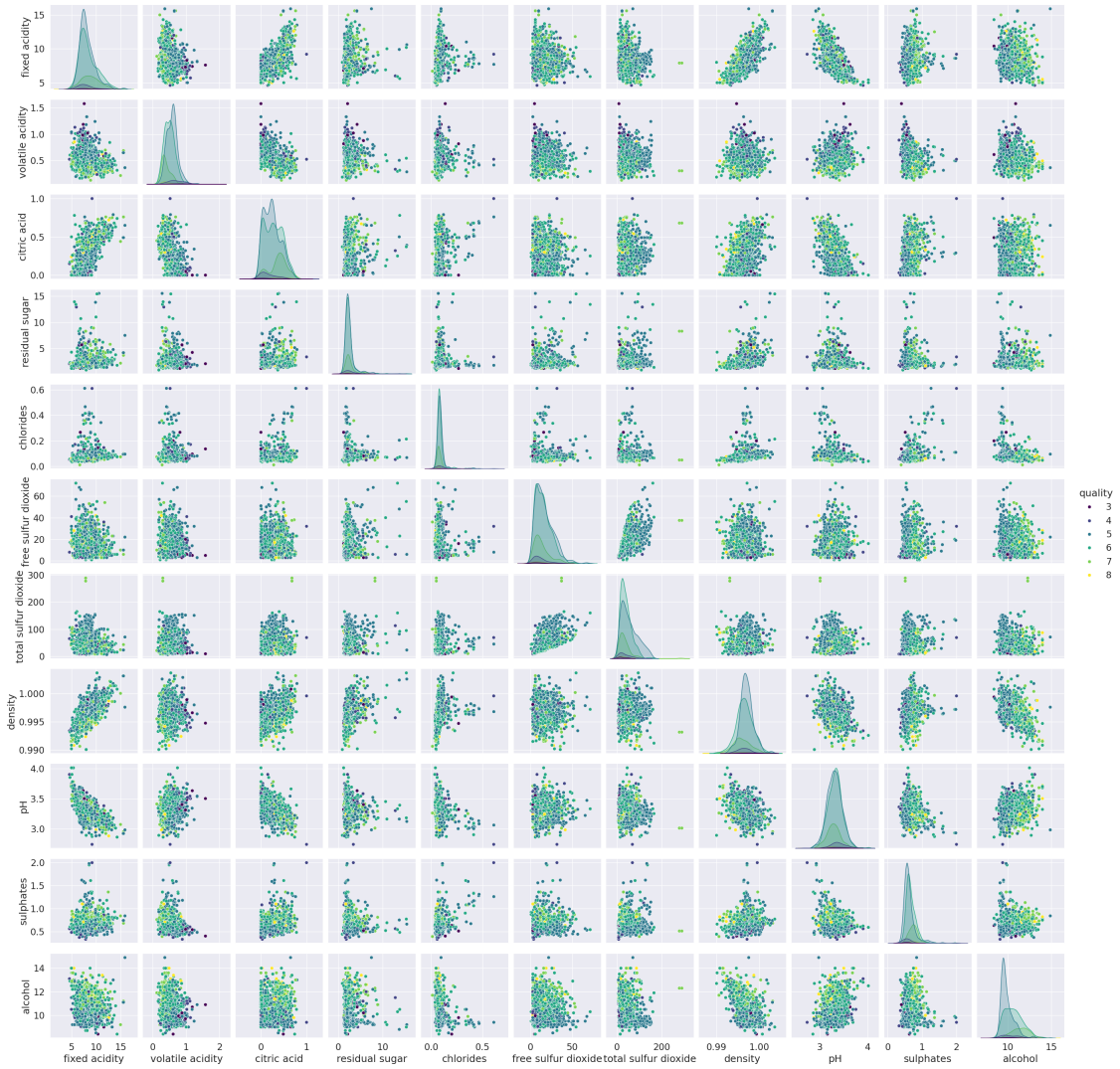

Boxplot of density by Quality
1.004
1.002
1.000
0.998
0.996
0.994
0.992
0.990
density
3
4
5
6
7
8
quality


Boxplot of pH by Quality
pH
4.0
3.8
3.6
3.4
3.2
3.0
2.8
3
4
5
6
7
8
quality


Boxplot of sulphates by Quality
sulphates
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
3
4
5
6
7
8
quality

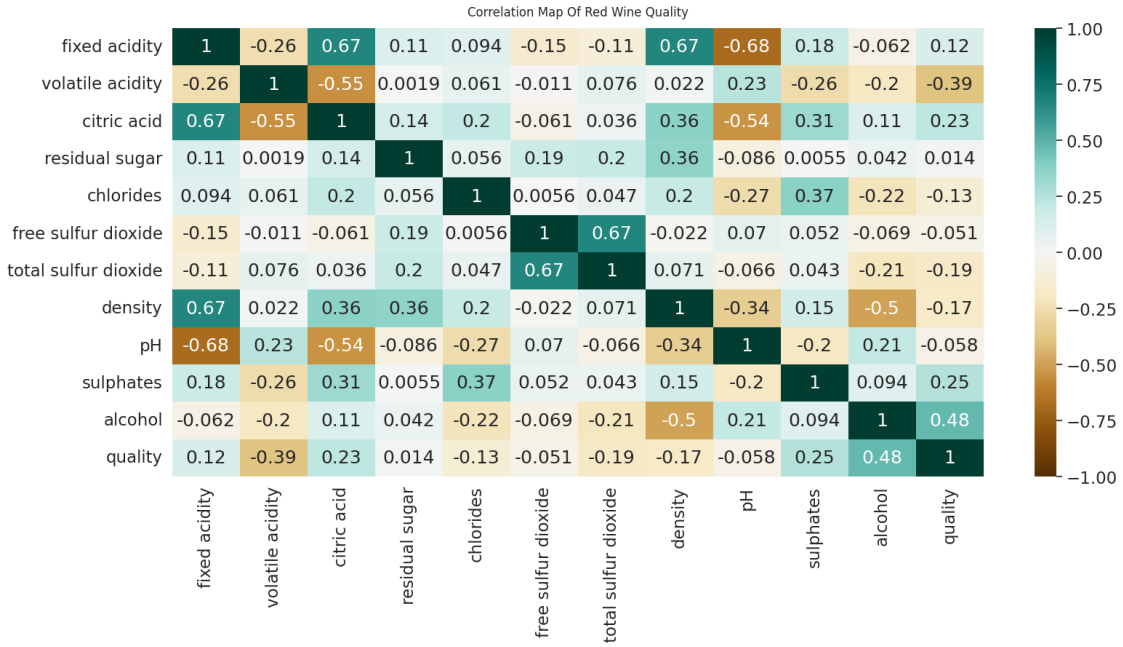


[4275]:
```
import seaborn as sns

selected_predictors = ['fixed acidity', 'volatile acidity', 'citric acid',␣
 ↪'residual sugar', 'chlorides', 'free sulfur dioxide', 'total sulfur␣
 ↪dioxide', 'density', 'pH', 'sulphates', 'alcohol', 'quality']

sns.pairplot(wine_data[selected_predictors], hue='quality', palette='viridis')
plt.show()
```



[4276]:
```
import plotly.express as px
import pickle
```

[4277]:
```
plt.figure(figsize=(18, 8))
sns.heatmap(wine_data.corr(), vmin=-1, vmax=1, annot=True, cmap='BrBG')
plt.title('Correlation Map Of Red Wine Quality', fontdict={'fontsize':12},␣
 ↪pad=12);
```


Correlation Map Of Red Wine Quality
fixed acidity 1 -0.26 0.67 0.11 0.094 -0.15 -0.11 0.67 -0.68 0.18 -0.062 0.12
volatile acidity -0.26 1 -0.55 0.0019 0.061 -0.011 0.076 0.022 0.23 -0.26 -0.2 -0.39
citric acid 0.67 -0.55 1 0.14 0.2 -0.061 0.036 0.36 -0.54 0.31 0.11 0.23
residual sugar 0.11 0.0019 0.14 1 0.056 0.19 0.2 0.36 -0.086 0.0055 0.042 0.014
chlorides 0.094 0.061 0.2 0.056 1 0.0056 0.047 0.2 -0.27 0.37 -0.22 -0.13
free sulfur dioxide -0.15 -0.011 -0.061 0.19 0.0056 1 0.67 -0.022 0.07 0.052 -0.069 -0.051
total sulfur dioxide -0.11 0.076 0.036 0.2 0.047 0.67 1 0.071 -0.066 0.043 -0.21 -0.19
density 0.67 0.022 0.36 0.36 0.2 -0.022 0.071 1 -0.34 0.15 -0.5 -0.17
pH -0.68 0.23 -0.54 -0.086 -0.27 0.07 -0.066 -0.34 1 -0.2 0.21 -0.058
sulphates 0.18 -0.26 0.31 0.0055 0.37 0.052 0.043 0.15 -0.2 1 0.094 0.25
alcohol -0.062 -0.2 0.11 0.042 -0.22 -0.069 -0.21 -0.5 0.21 0.094 1 0.48
quality 0.12 -0.39 0.23 0.014 -0.13 -0.051 -0.19 -0.17 -0.058 0.25 0.48 1
fixed acidity
volatile acidity
citric acid
residual sugar
chlorides
free sulfur dioxide
total sulfur dioxide
density
pH
sulphates
alcohol
quality
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00