In [1]:

```
import pandas as pd
```

In [2]:

```
import numpy as np
```

In [3]:

```
data=pd.read_csv("winequality-red.csv")
```

In [4]:

```
data.head()
```

Out[4]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcoh
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9

In [5]:

```
data.describe()
```

Out[5]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total su dio
count	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000
mean	8.319637	0.527821	0.270976	2.538806	0.087467	15.874922	46.467
std	1.741096	0.179060	0.194801	1.409928	0.047065	10.460157	32.895
min	4.600000	0.120000	0.000000	0.900000	0.012000	1.000000	6.000
25%	7.100000	0.390000	0.090000	1.900000	0.070000	7.000000	22.000
50%	7.900000	0.520000	0.260000	2.200000	0.079000	14.000000	38.000
75%	9.200000	0.640000	0.420000	2.600000	0.090000	21.000000	62.000
max	15.900000	1.580000	1.000000	15.500000	0.611000	72.000000	289.000

In [9]:

```
data.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 [8]:

```
data.isna().any()
```

Out[8]:

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

In [10]:

```
import seaborn as sns
import matplotlib.pyplot as plt
```

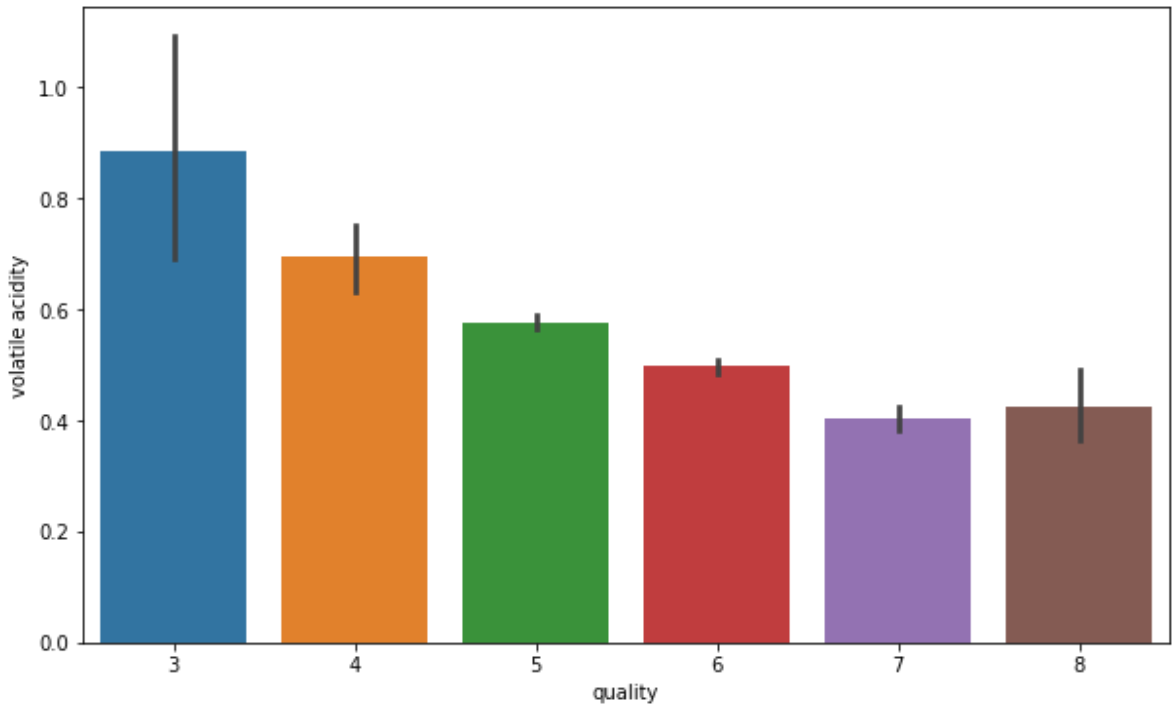
In [26]:

```
#Here we see that its quite a downing trend in the volatile acidity as we go higher the qua

fig = plt.figure(figsize = (10,6))
sns.barplot(x = 'quality', y = 'volatile acidity', data =data)
```

Out[26]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fa1eb7e42e8>
```

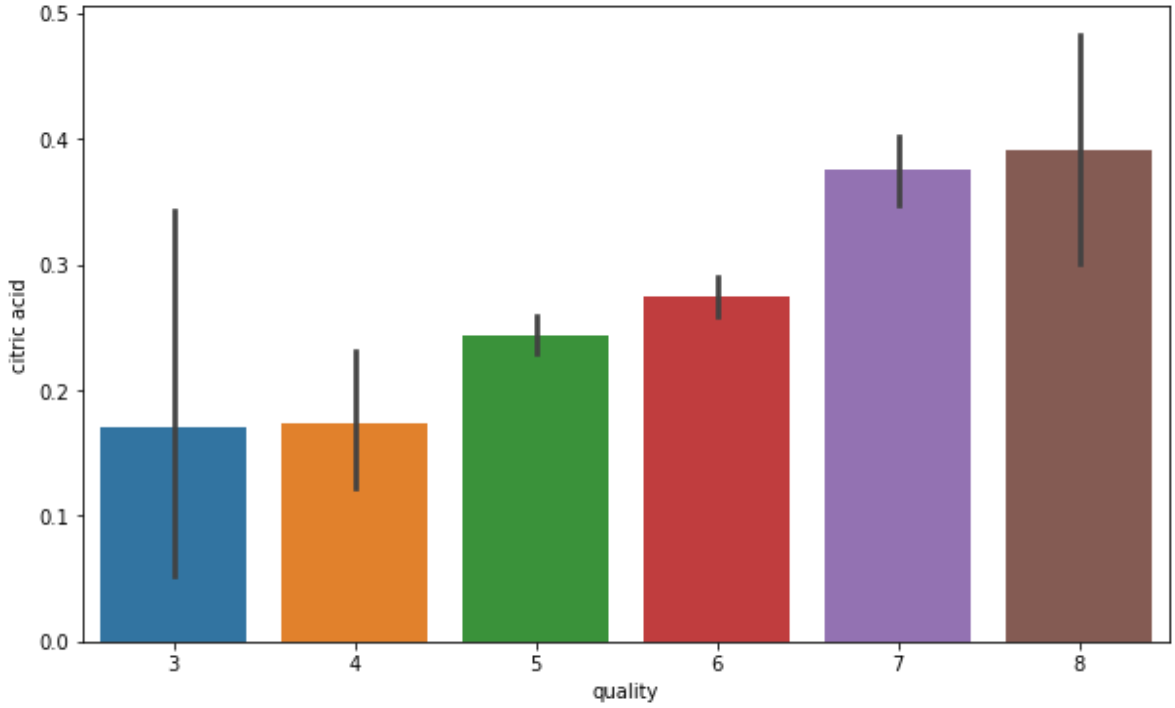


In [27]:

```
#Composition of citric acid go higher as we go higher in the quality of the wine
fig = plt.figure(figsize = (10,6))
sns.barplot(x = 'quality', y = 'citric acid', data = data)
```

Out[27]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fa1eb756b70>
```

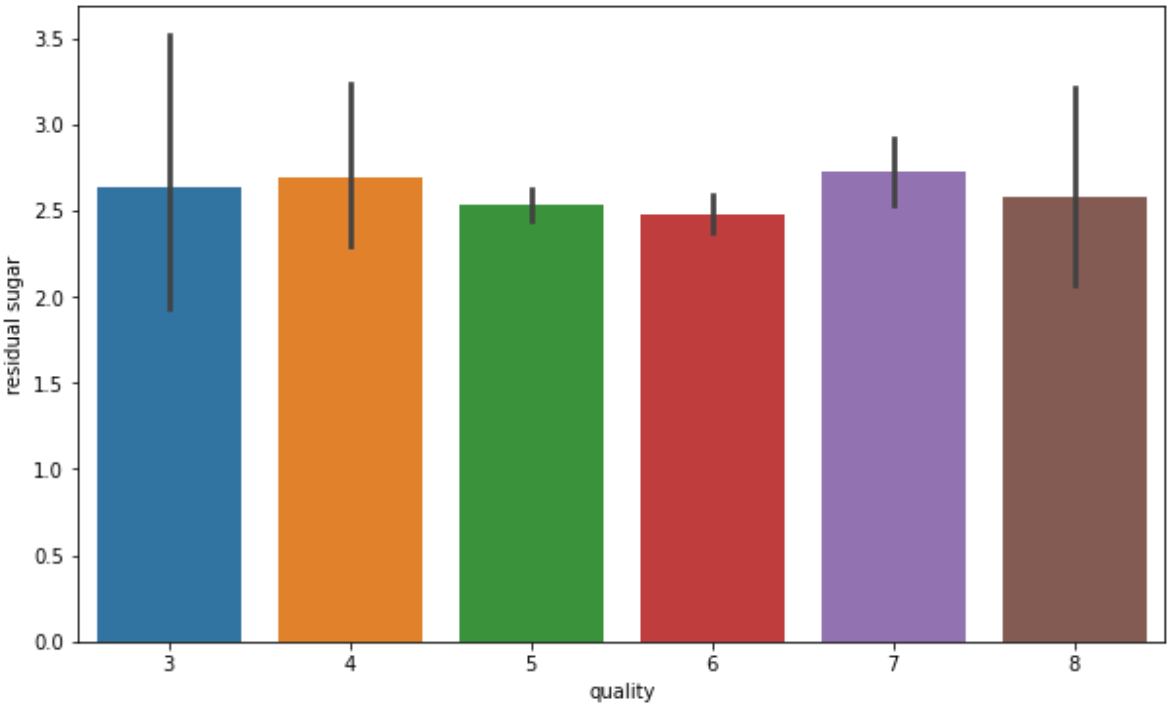


In [28]:

```
fig = plt.figure(figsize = (10,6))
sns.barplot(x = 'quality', y = 'residual sugar', data = data)
```

Out[28]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fa1eb735ba8>
```

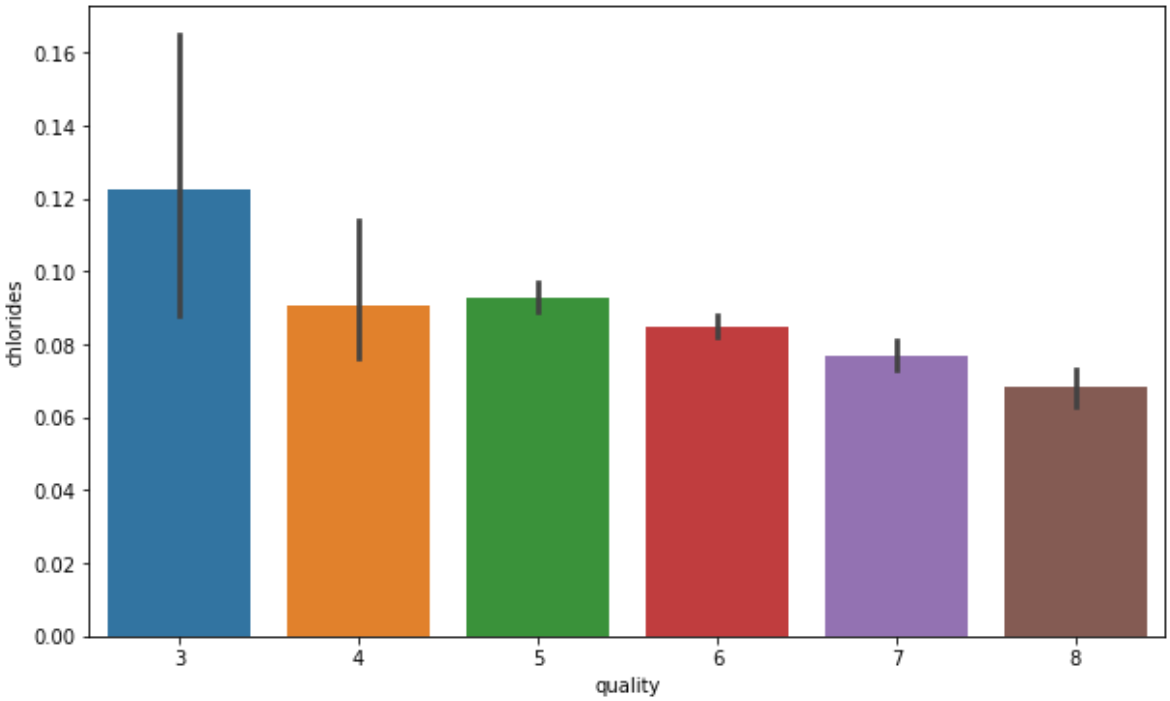


In [29]:

```
#Composition of chloride also go down as we go higher in the quality of the wine
fig = plt.figure(figsize = (10,6))
sns.barplot(x = 'quality', y = 'chlorides', data = data)
```

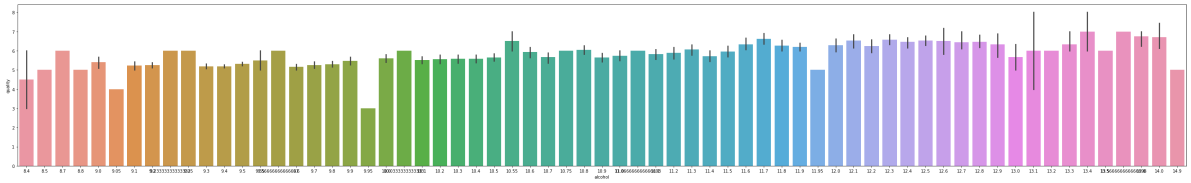
Out[29]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fa1eb647f98>
```



In [15]:

```
plt.figure(figsize=(50,7))
ax = sns.barplot(x="alcohol", y="quality", data=data)
```



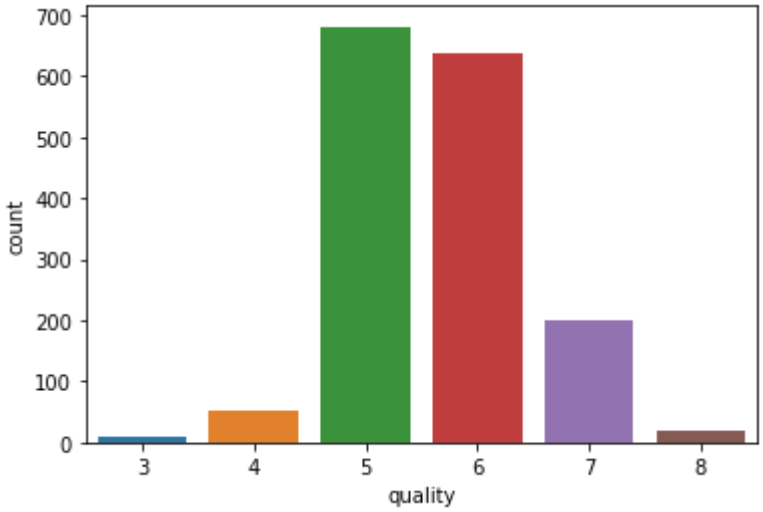
In [30]:

```
sns.countplot(data['quality'])
```

```
/usr/local/lib/python3.6/dist-packages/seaborn/_decorators.py:43: FutureWar
ing: Pass the following variable as a keyword arg: x. From version 0.12, the
only valid positional argument will be `data`, and passing other arguments w
ithout an explicit keyword will result in an error or misinterpretation.
  FutureWarning
```

Out[30]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fa1eb58c080>
```



In [16]:

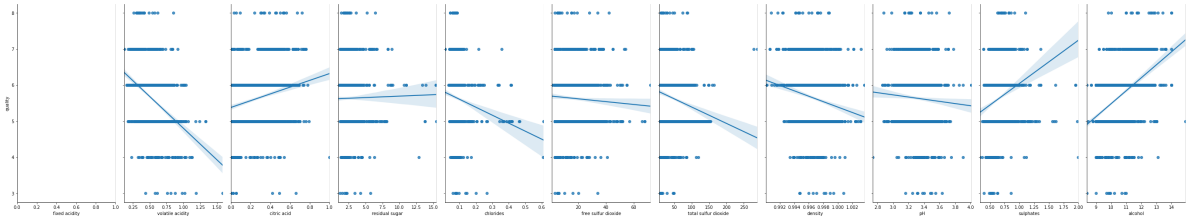
```
features = ['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar',
       'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'density',
       'pH', 'sulphates', 'alcohol']
x = data[features]
y = data['quality']

import seaborn as sns
#plotting features vs quality
sns.pairplot(data,x_vars=features,y_vars='quality',kind='reg',size=7,aspect=0.5)
```

```
/usr/local/lib/python3.6/dist-packages/seaborn/axisgrid.py:1912: UserWarnir
g: The `size` parameter has been renamed to `height`; please update your cod
e.
  warnings.warn(msg, UserWarning)
```

Out[16]:

```
<seaborn.axisgrid.PairGrid at 0x7fa1f30c3550>
```



In [21]:

```
sns.barplot(x='fixed acidity',y='quality',data=data)
```

Out[21]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fa1f3c64128>
```