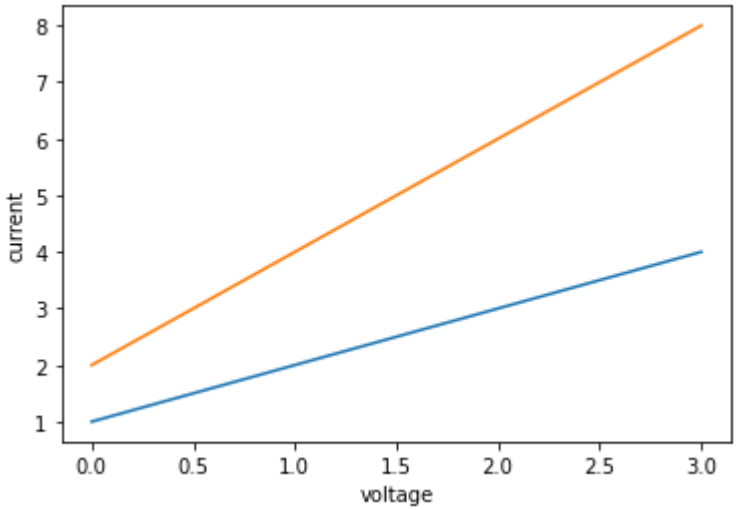
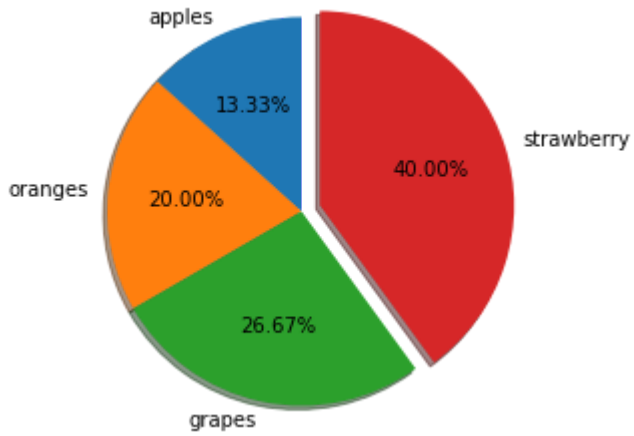
Data Visualization

In [1]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

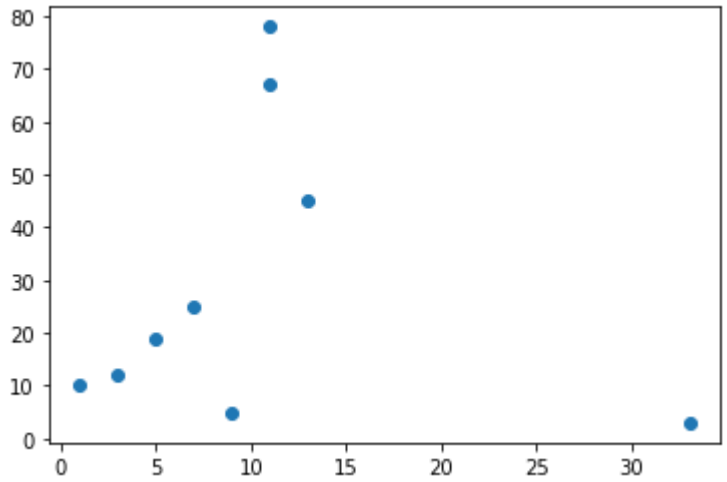
In [5]:
```
x=[1,2,3,4]
y=[2,4,6,8]
plt.plot(x)
plt.plot(y)
plt.ylabel('current')
plt.xlabel('voltage')
plt.show()
```



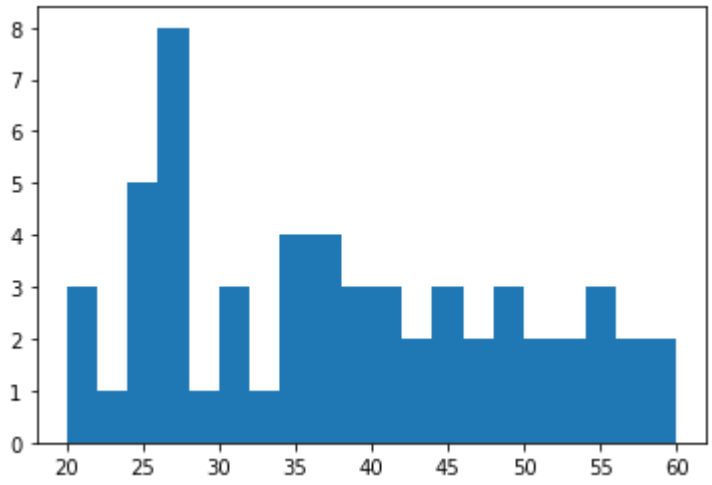
In [23]:
```
fruits='apples','oranges','grapes','strawberry'
sizes=[10,15,20,30]
ex=(0,0,0,0.1)

fig1,ax1=plt.subplots()
ax1.pie(sizes,explode=ex,labels=fruits,autopct='%1.2f%%',shadow=True,startangle=90)
ax1.axis('equal')
plt.show()
```



In [25]:
```
x=[1,3,5,7,9,11,13,33,11]
y=[10,12,19,25,5,67,45,3,78]
plt.scatter(x,y)
plt.show()
```



In [39]:
```
x=[20,21,21,23,24,24,25,25,25,26,26,26,27,27,27,27,27,29,30,30,31,33,34,34,34,35,36,36,
plt.hist(x,bins=20)
plt.show()
```



In [40]:
```
sns.set(color_codes=True)
```

In [42]:
```
data=pd.read_csv('C://Users//harsh//Desktop//minor spec//winequality-red.csv',sep=';',h
data.head(10)
```

Out[42]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quali
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	

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quali
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	
5	7.4	0.66	0.00	1.8	0.075	13.0	40.0	0.9978	3.51	0.56	9.4	
6	7.9	0.60	0.06	1.6	0.069	15.0	59.0	0.9964	3.30	0.46	9.4	
7	7.3	0.65	0.00	1.2	0.065	15.0	21.0	0.9946	3.39	0.47	10.0	
8	7.8	0.58	0.02	2.0	0.073	9.0	18.0	0.9968	3.36	0.57	9.5	
9	7.5	0.50	0.36	6.1	0.071	17.0	102.0	0.9978	3.35	0.80	10.5	

In [53]:
```
x=data.shape
y=data.info()
z=data.describe()
y
```

```
<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 [50]:
```
x
```

Out[50]: (1599, 12)

In [52]:
```
z
```

Out[52]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	
count	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599
mean	8.319637	0.527821	0.270976	2.538806	0.087467	15.874922	46.467792	(
std	1.741096	0.179060	0.194801	1.409928	0.047065	10.460157	32.895324	(

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	
min	4.600000	0.120000	0.000000	0.900000	0.012000	1.000000	6.000000	(
25%	7.100000	0.390000	0.090000	1.900000	0.070000	7.000000	22.000000	(
50%	7.900000	0.520000	0.260000	2.200000	0.079000	14.000000	38.000000	(
75%	9.200000	0.640000	0.420000	2.600000	0.090000	21.000000	62.000000	(
max	15.900000	1.580000	1.000000	15.500000	0.611000	72.000000	289.000000	1

In [54]:
```
data['pH'].is_unique
```

Out[54]:
```
False
```

In [57]:
```
data.pH.unique()
```

Out[57]:
```
array([3.51, 3.2 , 3.26, 3.16, 3.3 , 3.39, 3.36, 3.35, 3.28, 3.58, 3.17,
       3.11, 3.38, 3.04, 3.52, 3.43, 3.34, 3.47, 3.46, 3.45, 3.4 , 3.42,
       3.23, 3.5 , 3.33, 3.21, 3.48, 3.9 , 3.25, 3.32, 3.15, 3.41, 3.44,
       3.31, 3.54, 3.13, 2.93, 3.14, 3.75, 3.85, 3.29, 3.08, 3.37, 3.19,
       3.07, 3.49, 3.53, 3.24, 3.63, 3.22, 3.68, 2.74, 3.59, 3.  , 3.12,
       3.57, 3.61, 3.06, 3.6 , 3.69, 3.1 , 3.05, 3.67, 3.27, 3.18, 3.02,
       3.55, 2.99, 3.01, 3.56, 3.03, 3.62, 2.88, 2.95, 2.98, 3.09, 2.86,
       3.74, 2.92, 3.72, 2.87, 2.89, 2.94, 3.66, 3.71, 3.78, 3.7 , 4.01,
       2.9 ])
```

In [58]:
```
data.pH.value_counts()
```

Out[58]:
```
3.30    57
3.36    56
3.26    53
3.38    48
3.39    48
        ..
2.86     1
3.75     1
2.74     1
2.90     1
3.74     1
Name: pH, Length: 89, dtype: int64
```

In [59]:
```
data.dtypes
```

Out[59]:
```
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
dtype: object
```

In [60]:
```
dupli=data[data.duplicated()]
print(f'duplicate rows are:{dupli.shape}')
```

```
duplicate rows are:(240, 12)
```

In [64]:
```
new_data=data.drop_duplicates()
new_data.head()
```

Out[64]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quali
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	
5	7.4	0.66	0.00	1.8	0.075	13.0	40.0	0.9978	3.51	0.56	9.4	

In [65]:
```
new_data.shape
```

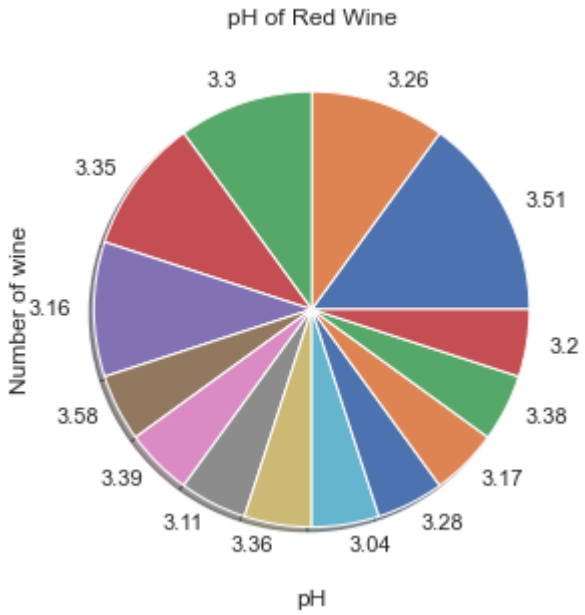
Out[65]:
```
(1359, 12)
```

Graphical Analysis

In [79]:
```
q=data.head(20)
```

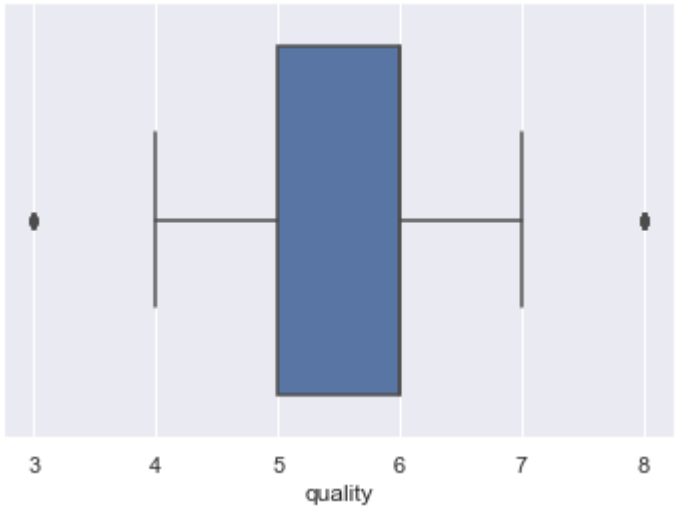
In [78]:
```
q.pH.value_counts().plot(kind='pie',figsize=(10,5),shadow=True)
plt.title('pH of Red Wine')
plt.ylabel('Number of wine')
plt.xlabel('pH')
```

Out[78]:
```
Text(0.5, 0, 'pH')
```



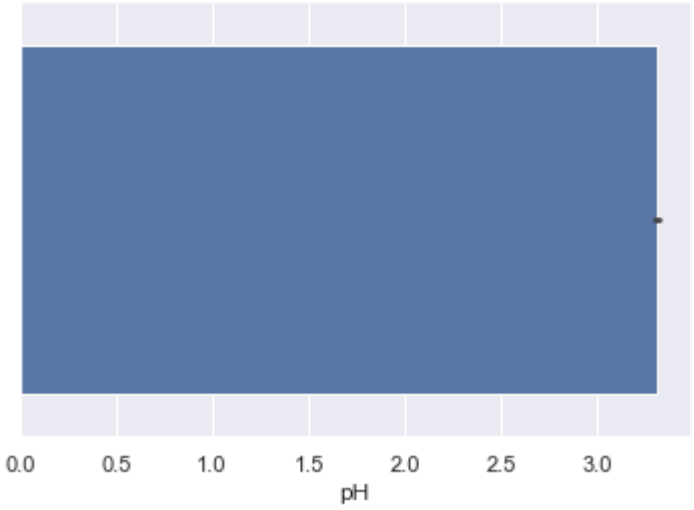
In [80]:
```
sns.boxplot(x=data['quality'])
```

Out[80]: <AxesSubplot:xlabel='quality'>



In [85]:
```
sns.barplot(x=data['pH'])
```

Out[85]: <AxesSubplot:xlabel='pH'>

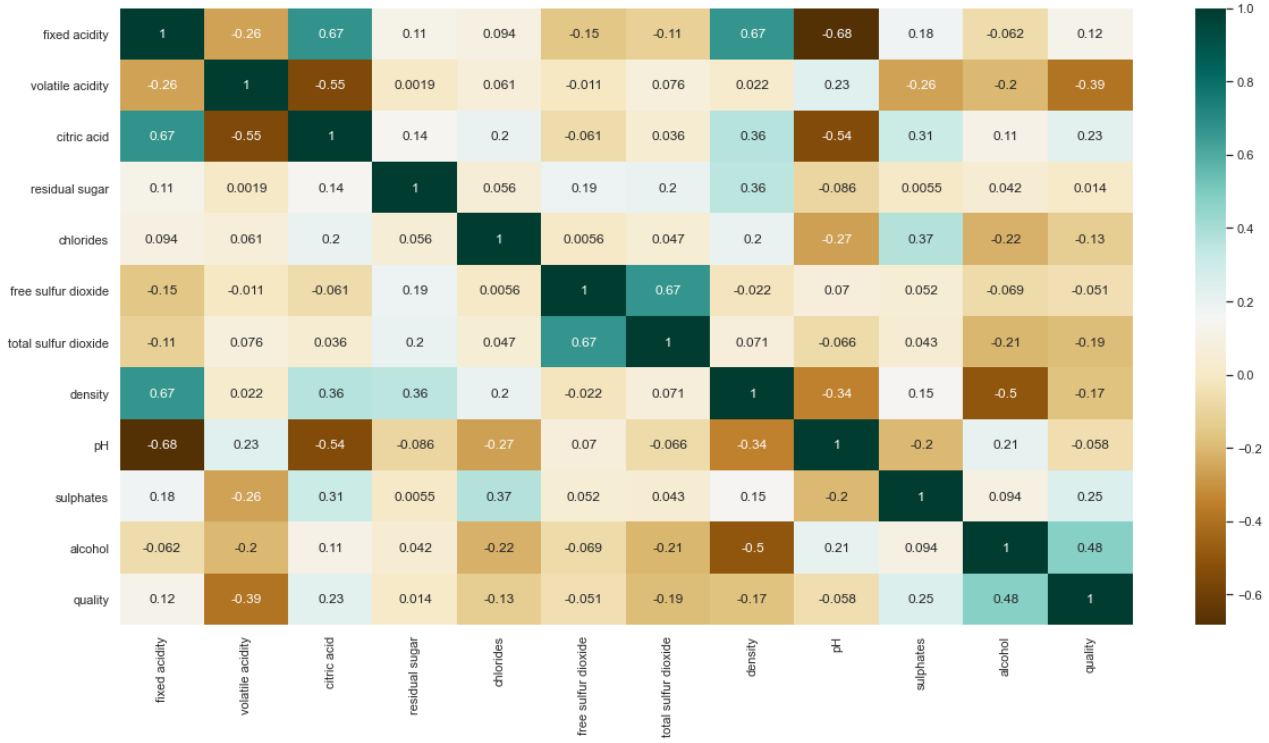


In [93]:

```
plt.figure(figsize=(20,10))
w=data.corr()
sns.heatmap(w,cmap='BrBG',annot=True)
w
```

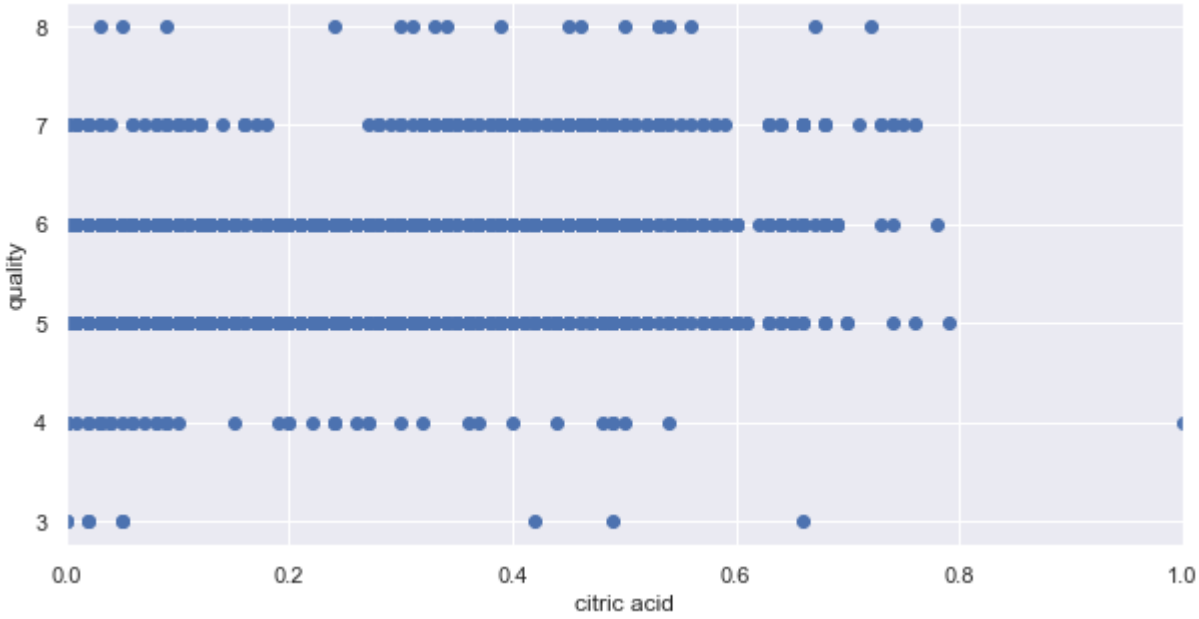
Out[93]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	p
fixed acidity	1.000000	-0.256131	0.671703	0.114777	0.093705	-0.153794	-0.113181	0.668047	-0.6829
volatile acidity	-0.256131	1.000000	-0.552496	0.001918	0.061298	-0.010504	0.076470	0.022026	0.2349
citric acid	0.671703	-0.552496	1.000000	0.143577	0.203823	-0.060978	0.035533	0.364947	-0.5419
residual sugar	0.114777	0.001918	0.143577	1.000000	0.055610	0.187049	0.203028	0.355283	-0.0856
chlorides	0.093705	0.061298	0.203823	0.055610	1.000000	0.005562	0.047400	0.200632	-0.2650
free sulfur dioxide	-0.153794	-0.010504	-0.060978	0.187049	0.005562	1.000000	0.667666	-0.021946	0.0703
total sulfur dioxide	-0.113181	0.076470	0.035533	0.203028	0.047400	0.667666	1.000000	0.071269	-0.0664
density	0.668047	0.022026	0.364947	0.355283	0.200632	-0.021946	0.071269	1.000000	-0.3416
pH	-0.682978	0.234937	-0.541904	-0.085652	-0.265026	0.070377	-0.066495	-0.341699	1.0000
sulphates	0.183006	-0.260987	0.312770	0.005527	0.371260	0.051658	0.042947	0.148506	-0.1966
alcohol	-0.061668	-0.202288	0.109903	0.042075	-0.221141	-0.069408	-0.205654	-0.496180	0.2056
quality	0.124052	-0.390558	0.226373	0.013732	-0.128907	-0.050656	-0.185100	-0.174919	-0.0577



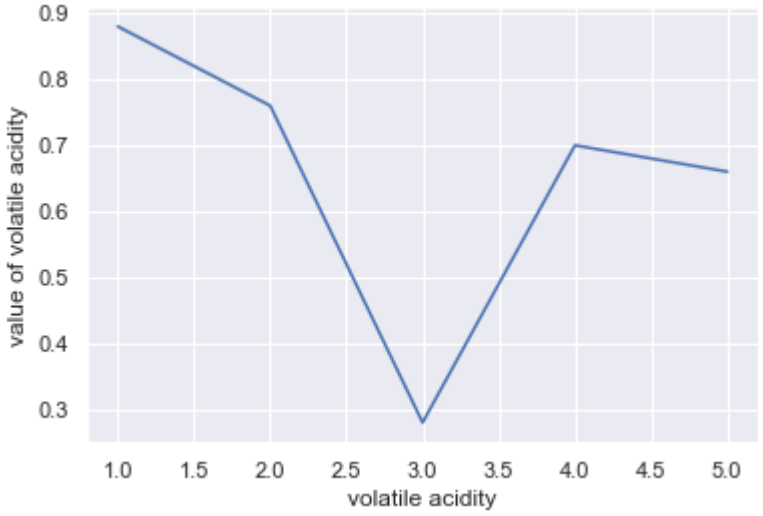
In [96]:

```
fig,ax=plt.subplots(figsize=(10,5))
plt.xlim((0.0,1.0))
plt.scatter(data['citric acid'],data['quality'])
plt.xlabel('citric acid')
plt.ylabel('quality')
plt.show()
```



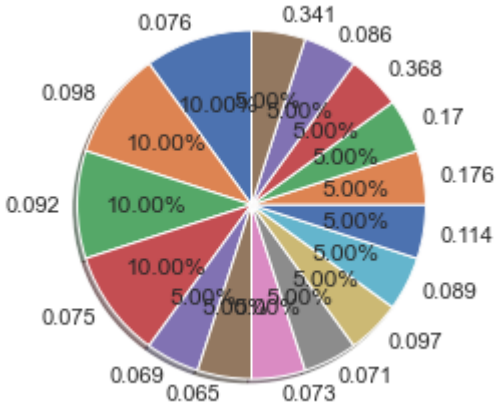
In [101…

```
plt.plot(data['volatile acidity'][1:6])
plt.ylabel('value of volatile acidity')
plt.xlabel('volatile acidity')
plt.show()
```



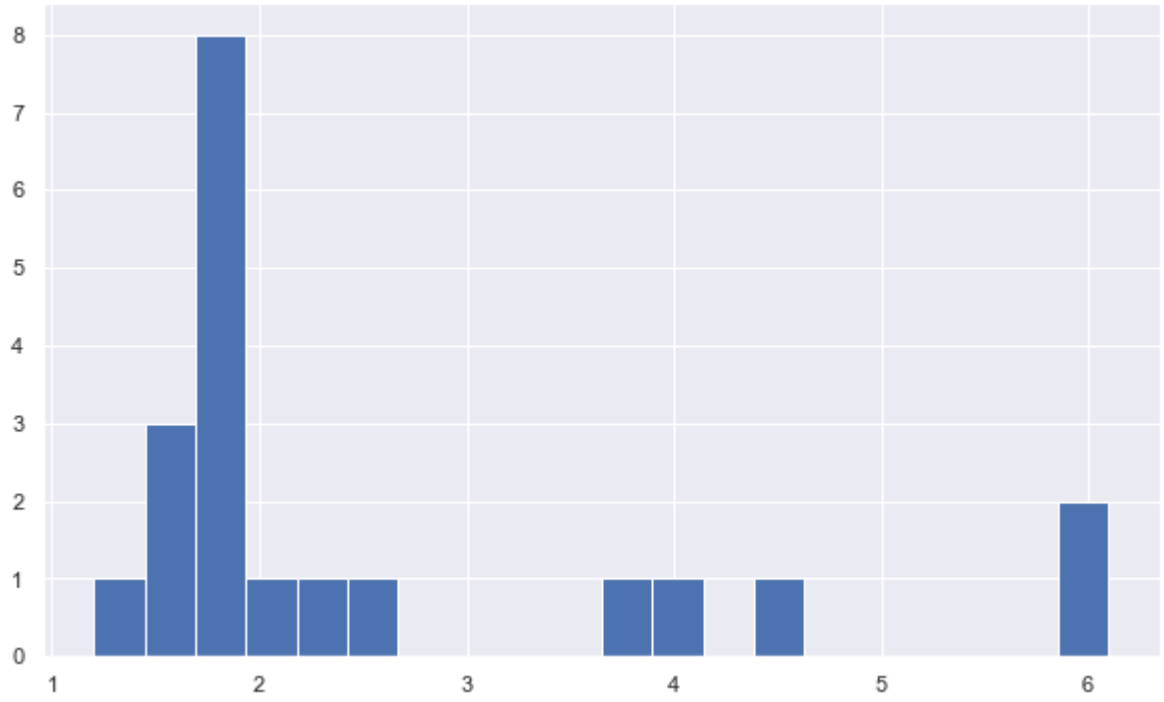
In [109…

```
lab=q.chlorides.unique()
sizes=q.chlorides.value_counts()
fig1,ax=plt.subplots()
ax.pie(sizes,labels=lab,autopct='%1.2f%%',shadow=True,startangle=90)
plt.show()
```



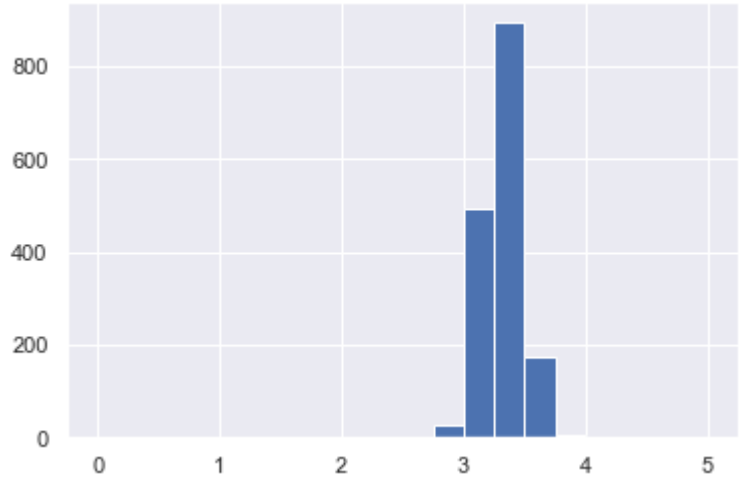
In [111…

```
fig,ax2=plt.subplots(figsize=(10,6))
ax2.hist(q['residual sugar'],bins=20)
plt.show()
```



In [118…

```
plt.hist(data['pH'],bins=20,range=[0,5.0])
plt.show()
```



In [ ]: