```
import pandas as pd
import numpy as np


file_name = "/content/winequalityN.csv"
df = pd.read_csv(file_name)
df.head()
```

	type	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	white	7.0	0.27	0.36	20.7	0.045	45.0	170.0	1.0010	3.00	0.45	8.8	6
1	white	6.3	0.30	0.34	1.6	0.049	14.0	132.0	0.9940	3.30	0.49	9.5	6
2	white	8.1	0.28	0.40	6.9	0.050	30.0	97.0	0.9951	3.26	0.44	10.1	6
3	white	7.2	0.23	0.32	8.5	0.058	47.0	186.0	0.9956	3.19	0.40	9.9	6
4	white	7.2	0.23	0.32	8.5	0.058	47.0	186.0	0.9956	3.19	0.40	9.9	6

```
Start coding or generate with AI.
```

```
#simple descriptive analysis
df.describe()
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	s di
count	6487.000000	6489.000000	6494.000000	6495.000000	6495.000000	6497.000000	6497.0
mean	7.216579	0.339691	0.318722	5.444326	0.056042	30.525319	115.7
std	1.296750	0.164649	0.145265	4.758125	0.035036	17.749400	56.5
min	3.800000	0.080000	0.000000	0.600000	0.009000	1.000000	6.0
25%	6.400000	0.230000	0.250000	1.800000	0.038000	17.000000	77.0
50%	7.000000	0.290000	0.310000	3.000000	0.047000	29.000000	118.0
75%	7.700000	0.400000	0.390000	8.100000	0.065000	41.000000	156.0
max	15.900000	1.580000	1.660000	65.800000	0.611000	289.000000	440.0

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6497 entries, 0 to 6496
Data columns (total 13 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   type                  6497 non-null   object
 1   fixed acidity         6487 non-null   float64
 2   volatile acidity      6489 non-null   float64
 3   citric acid           6494 non-null   float64
 4   residual sugar        6495 non-null   float64
 5   chlorides             6495 non-null   float64
 6   free sulfur dioxide   6497 non-null   float64
 7   total sulfur dioxide  6497 non-null   float64
 8   density               6497 non-null   float64
 9   pH                    6488 non-null   float64
 10  sulphates             6493 non-null   float64
 11  alcohol               6497 non-null   float64
 12  quality               6497 non-null   int64
dtypes: float64(11), int64(1), object(1)
memory usage: 660.0+ KB
```

```
df["pH"].mean()
df["pH"].median()
df["pH"].mode()
```

```
0    3.16
Name: pH, dtype: float64
```

```
df["pH"].mode()
```

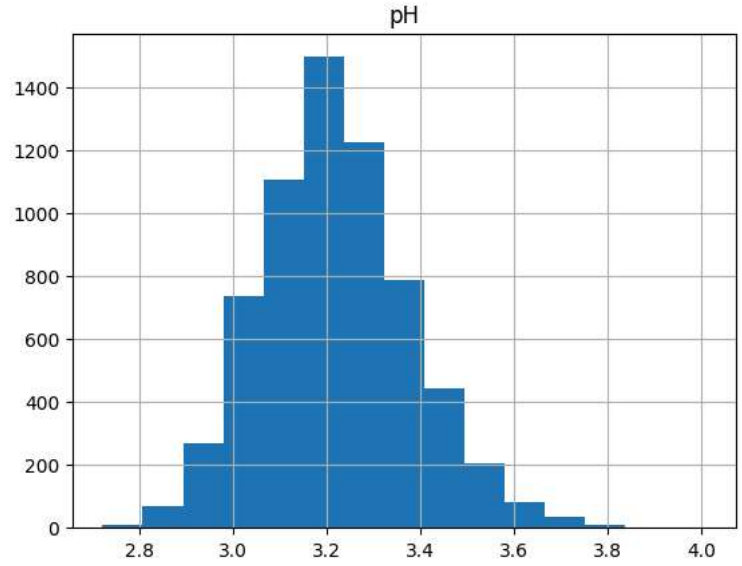
```
0    3.16
Name: pH, dtype: float64
```

```
df["pH"].median()
```

```
3.21
```

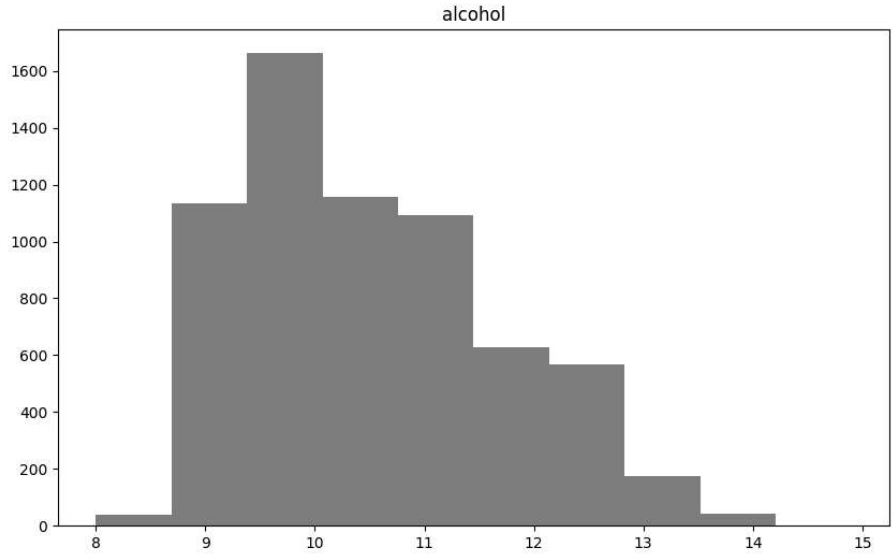
```
#lets start plotting using a histogram
df.hist(column="pH",bins=15)
```

```
array([[<Axes: title={'center': 'pH'}>]], dtype=object)
```



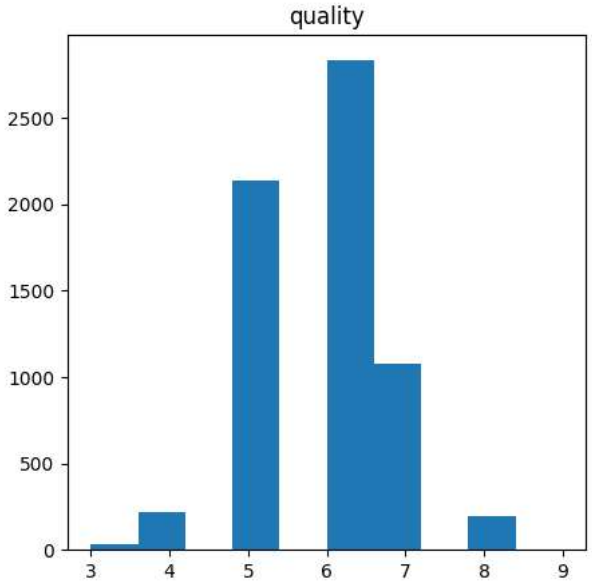

```
df.hist(column="alcohol",bins=10,grid=False,figsize=(10,6),color="grey")
```

```
array([[<Axes: title={'center': 'alcohol'}>]], dtype=object)
```



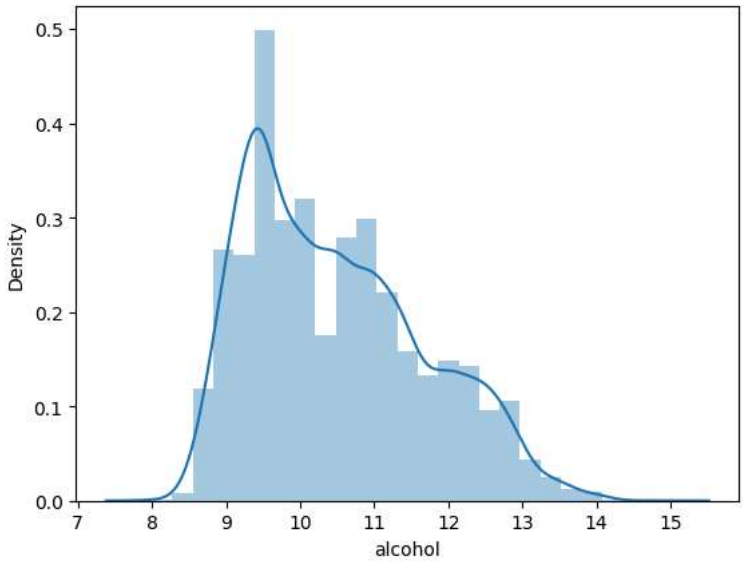

```
df.hist(column="quality",bins=10,grid=False,figsize=(5,5))
```

```
array([[<Axes: title={'center': 'quality'}>]], dtype=object)
```



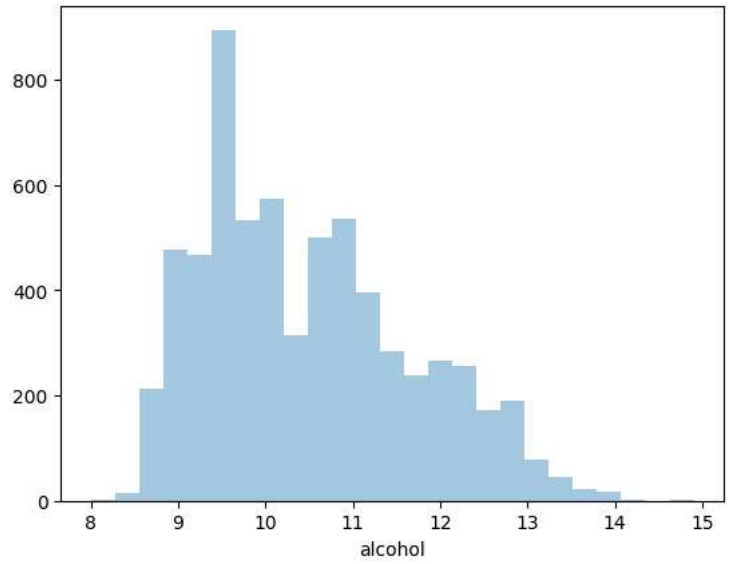

```
#lets tryout with library seaborn
import seaborn as sns
sns.distplot(df["alcohol"],bins=25,kde=True)
```

```
<ipython-input-6-74fbfada42b8>:3: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df["alcohol"],bins=25,kde=True)
<Axes: xlabel='alcohol', ylabel='Density'>
```



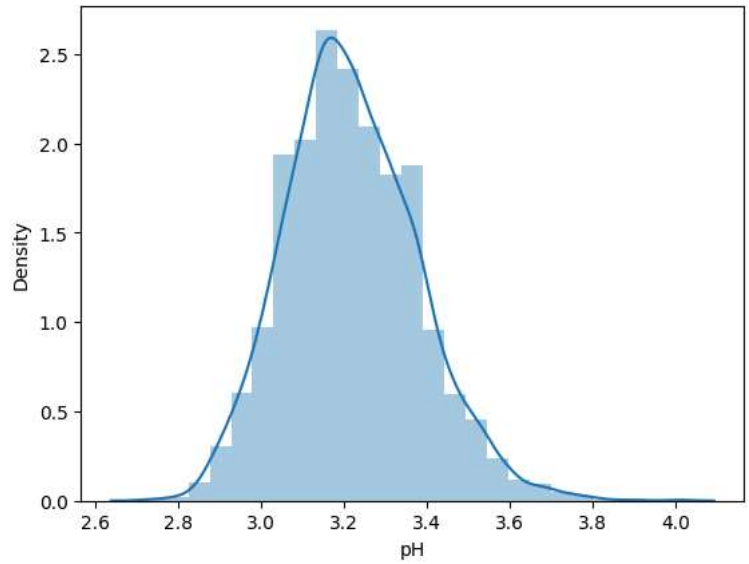

```
sns.distplot(df["alcohol"],bins=25,kde=False)
```

```
<ipython-input-19-379a9c9047a3>:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df["alcohol"],bins=25,kde=False)
<Axes: xlabel='alcohol'>
```




```
sns.distplot(df["pH"],bins=25,kde=True)
```

```
<ipython-input-20-4c6e4fd1f6f6>:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df["pH"],bins=25,kde=True)
<Axes: xlabel='pH', ylabel='Density'>
```




```
df["alcohol"].value_counts().head()
```

```
9.5     367
9.4     332
```

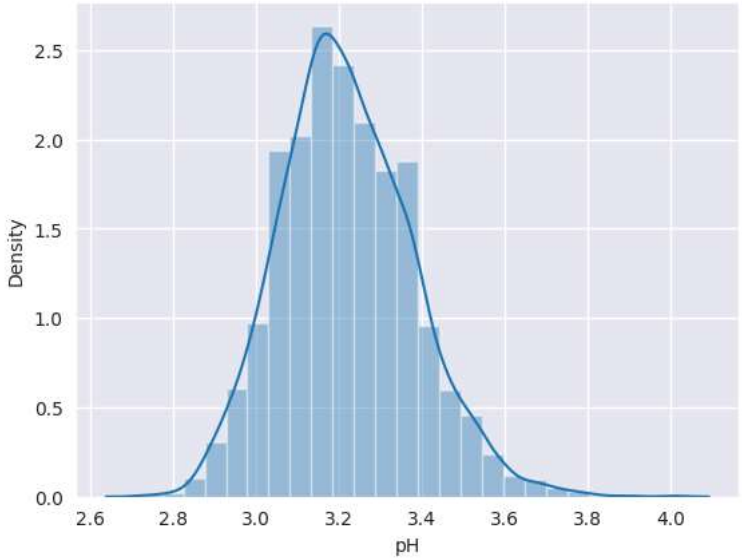
```
9.2     271
10.0    229
10.5    227
Name: alcohol, dtype: int64
```

```
grouped = df.groupby('alcohol')
print(grouped.get_group(9.5))
```

```
       type  fixed acidity  volatile acidity  citric acid  residual sugar  \
1     white            6.3             0.300         0.34             1.6
8     white            6.3             0.300         0.34             1.6
19    white            6.5             0.310         0.14             7.5
43    white            6.6             0.240         0.27             1.4
44    white            6.7             0.230         0.26             1.4
...     ...            ...               ...          ...             ...
6282    red            7.1             0.755         0.15             1.8
6448    red            7.1             0.680         0.00             2.3
6456    red            6.9             0.630         0.33             6.7
6470    red            7.3             0.690         0.32             2.2
6491    red            6.8             0.620         0.08             1.9

      chlorides  free sulfur dioxide  total sulfur dioxide  density    pH  \
1         0.049                 14.0                 132.0  0.99400  3.30
8         0.049                 14.0                 132.0  0.99400  3.30
19        0.044                 34.0                 133.0  0.99550  3.22
43        0.057                 33.0                 152.0  0.99340  3.22
44        0.060                 33.0                 154.0  0.99340  3.24
...         ...                  ...                   ...      ...   ...
6282      0.107                 20.0                  84.0  0.99593  3.19
6448      0.087                 17.0                  26.0  0.99783  3.45
6456      0.235                 66.0                 115.0  0.99787  3.22
6470      0.069                 35.0                 104.0  0.99632  3.33
6491      0.068                 28.0                  38.0  0.99651  3.42

      sulphates  alcohol  quality
1          0.49      9.5        6
8          0.49      9.5        6
19         0.50      9.5        5
43         0.56      9.5        6
44         0.56      9.5        6
...         ...      ...      ...
6282       0.50      9.5        5
6448       0.53      9.5        5
6456       0.56      9.5        5
6470       0.51      9.5        5
6491       0.82      9.5        6

[367 rows x 13 columns]
```

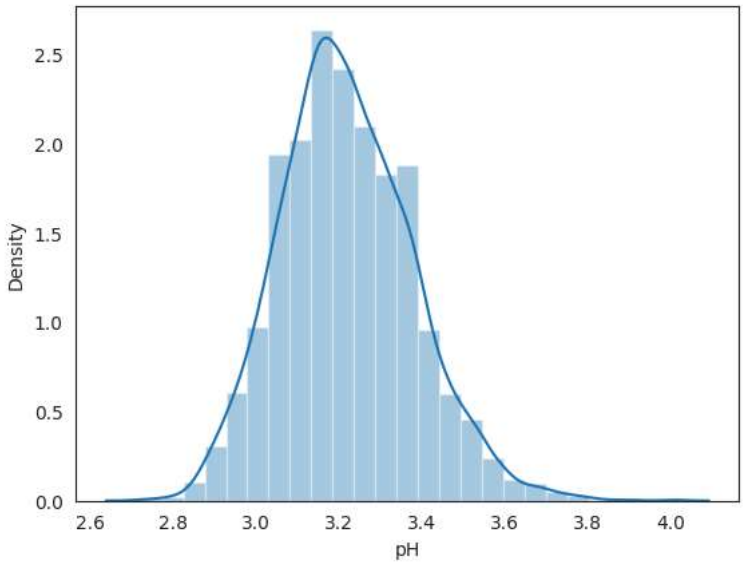
```
sns.set_style("darkgrid")
sns.distplot(df["pH"],bins=25,kde=True)
```

```
<ipython-input-36-28b6a84cef71>:2: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df["pH"],bins=25,kde=True)
<Axes: xlabel='pH', ylabel='Density'>
```



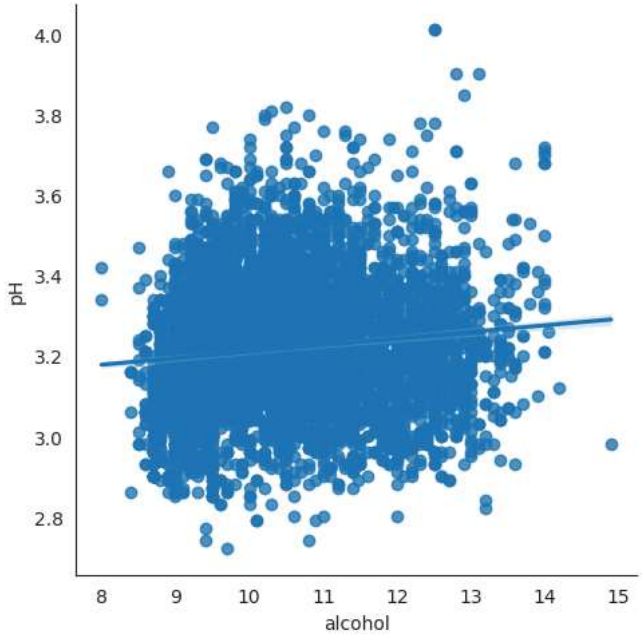

```
sns.set_style("white")
sns.distplot(df["pH"],bins=25,kde=True)
```

```
<ipython-input-37-6522619d4d72>:2: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df["pH"],bins=25,kde=True)
<Axes: xlabel='pH', ylabel='Density'>
```



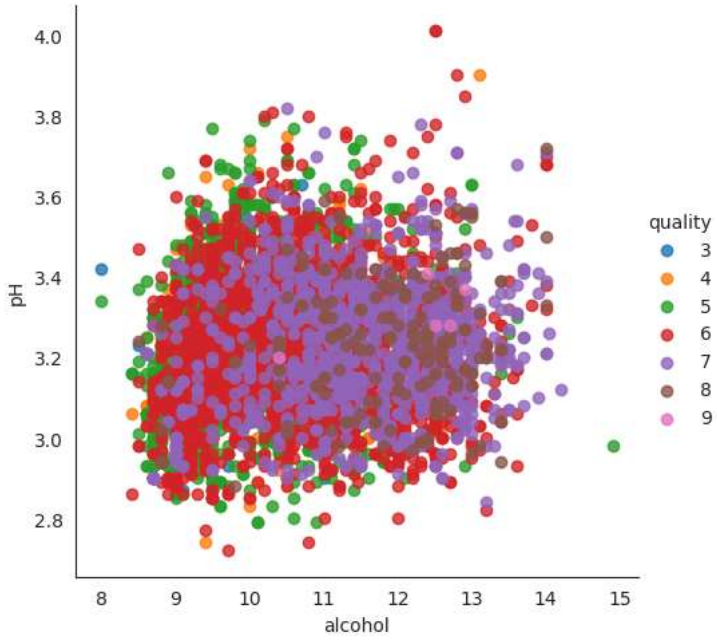

```
sns.lmplot(x="alcohol",y="pH",data=df)
```

```
<seaborn.axisgrid.FacetGrid at 0x7c003aac90c0>
```




```
sns.lmplot(x="alcohol",y="pH",data=df,fit_reg=False,hue="quality")
```

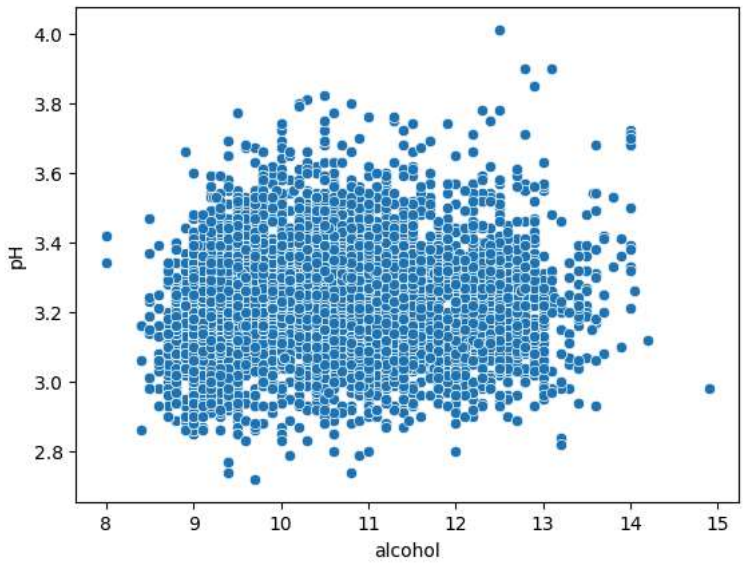
```
<seaborn.axisgrid.FacetGrid at 0x7c0029d0e3b0>
```



Start coding or generate with AI.

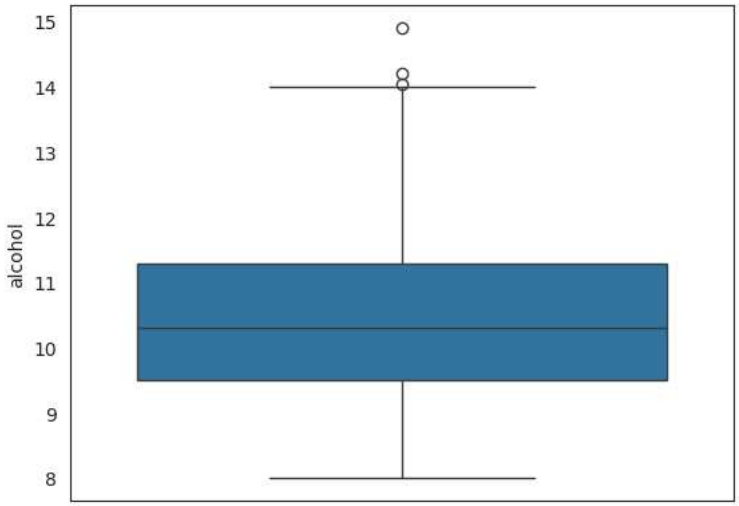
```
sns.scatterplot(x="alcohol",y="pH",data=df)
```

```
<Axes: xlabel='alcohol', ylabel='pH'>
```



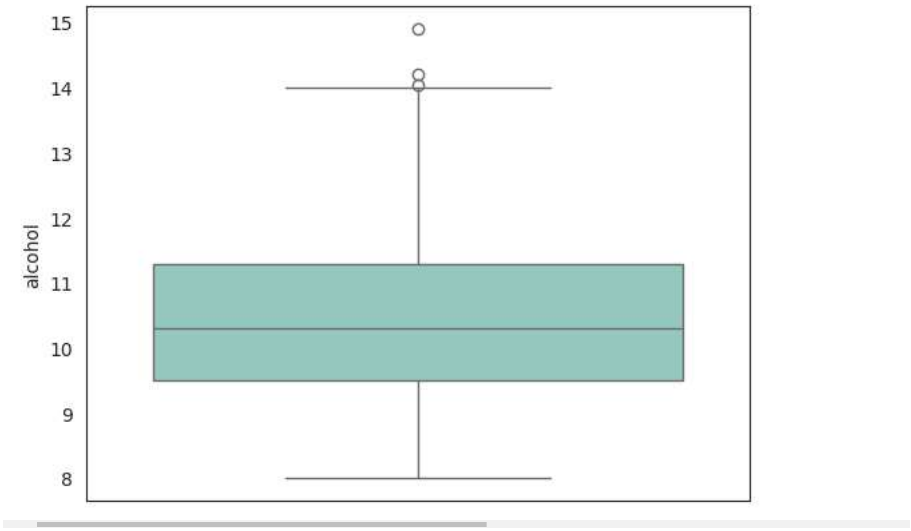

```
sns.boxplot(data=df["alcohol"])
```

```
<Axes: ylabel='alcohol'>
```



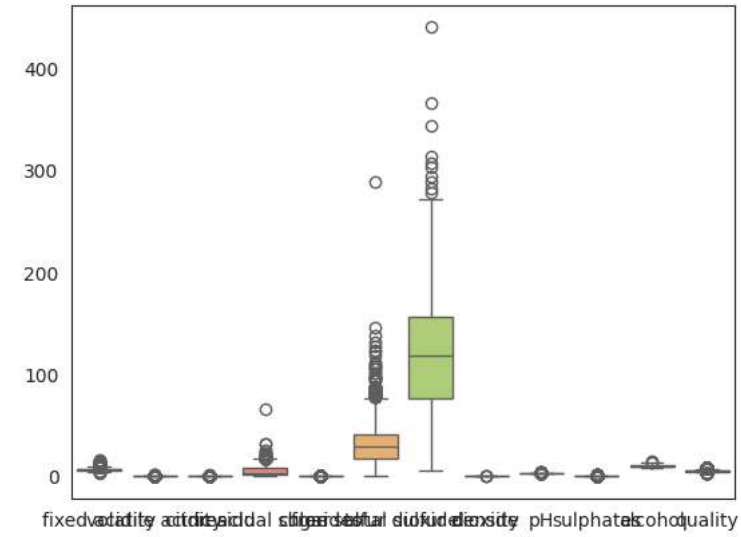

```
sns.boxplot(data=df["alcohol"],palette="Set3")
```

```
<ipython-input-47-6ac87147141a>:1: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0.

  sns.boxplot(data=df["alcohol"],palette="Set3")
<Axes: ylabel='alcohol'>
```




```
sns.boxplot(data=df,palette="Set3")
```

```
<Axes: >
```




```
sns.violinplot(data=df["alcohol"])
```

```
<Axes: ylabel='alcohol'>
```