

```
In [1]: import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import scipy.stats as st
import seaborn as sns
%matplotlib inline
```

```
In [2]: import warnings
warnings.filterwarnings('ignore')
```

```
In [3]: df= pd.read_csv(r"C:\Users\Jan Saida\OneDrive\Documents\heart.csv")
```

```
In [4]: df
```

```
Out[4]:
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

303 rows × 14 columns

```
In [5]: print('The shape of the dataset:',df.shape)
```

The shape of the dataset: (303, 14)

```
In [6]: df.shape
```

```
Out[6]: (303, 14)
```

```
In [7]: df.head()
```

```
Out[7]:
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

```
In [8]: df.tail()
```

```
Out[8]:
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

```
In [9]: df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 14 columns):
#   Column      Non-Null Count  Dtype
---  -
0    age         303 non-null    int64
1    sex         303 non-null    int64
2    cp          303 non-null    int64
3    trestbps    303 non-null    int64
4    chol        303 non-null    int64
5    fbs         303 non-null    int64
6    restecg     303 non-null    int64
7    thalach     303 non-null    int64
8    exang       303 non-null    int64
9    oldpeak     303 non-null    float64
10   slope       303 non-null    int64
11   ca          303 non-null    int64
12   thal        303 non-null    int64
13   target      303 non-null    int64
dtypes: float64(1), int64(13)
memory usage: 33.3 KB

```

```
In [10]: df.describe()
```

```
Out[10]:
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000
mean	54.366337	0.683168	0.966997	131.623762	246.264026	0.148515	0.528053	149.646865	0.326733	1.039604
std	9.082101	0.466011	1.032052	17.538143	51.830751	0.356198	0.525860	22.905161	0.469794	1.161075
min	29.000000	0.000000	0.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.000000
25%	47.500000	0.000000	0.000000	120.000000	211.000000	0.000000	0.000000	133.500000	0.000000	0.000000
50%	55.000000	1.000000	1.000000	130.000000	240.000000	0.000000	1.000000	153.000000	0.000000	0.800000
75%	61.000000	1.000000	2.000000	140.000000	274.500000	0.000000	1.000000	166.000000	1.000000	1.600000
max	77.000000	1.000000	3.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.200000

```
In [11]: df.dtypes
```

```
Out[11]: age          int64
sex          int64
cp           int64
trestbps     int64
chol         int64
fbs          int64
restecg      int64
thalach      int64
exang        int64
oldpeak      float64
slope        int64
ca           int64
thal         int64
target       int64
dtype: object
```

```
In [12]: df.columns
```

```
Out[12]: Index(['age', 'sex', 'cp', 'trestbps', 'chol', 'fbs', 'restecg', 'thalach',
               'exang', 'oldpeak', 'slope', 'ca', 'thal', 'target'],
              dtype='object')
```

```
In [13]: df['target'].nunique()
```

```
Out[13]: 2
```

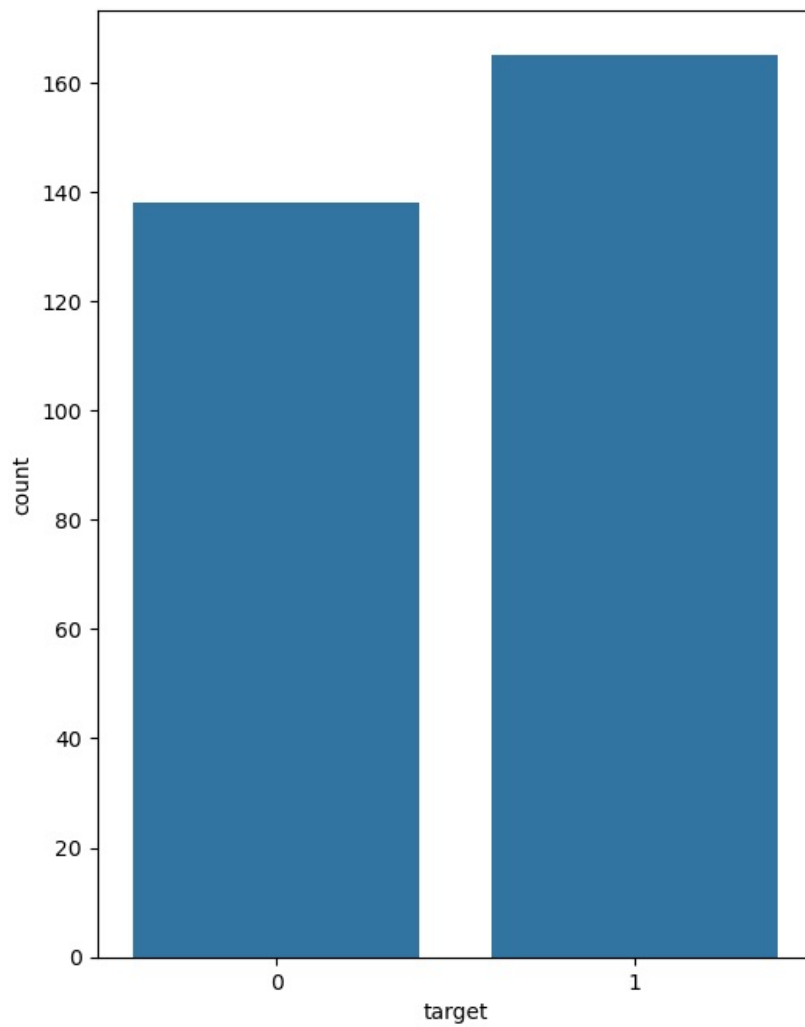
```
In [14]: df['target'].unique()
```

```
Out[14]: array([1, 0], dtype=int64)
```

```
In [15]: df['target'].value_counts()
```

```
Out[15]: target
1      165
0      138
Name: count, dtype: int64
```

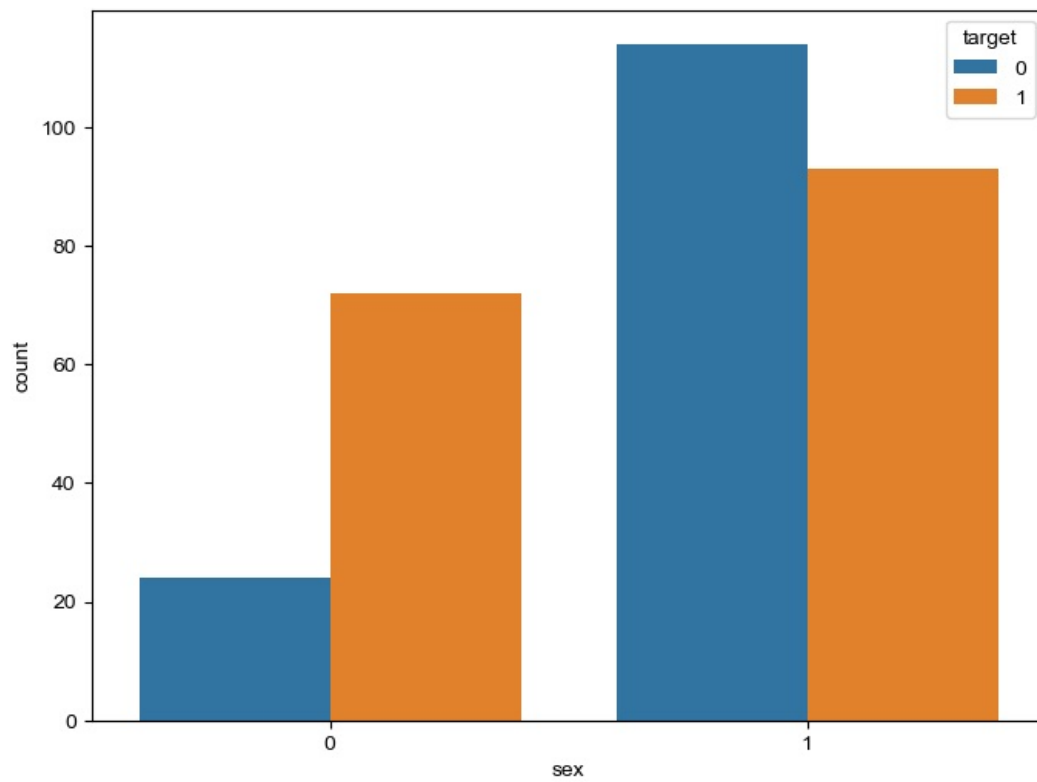
```
In [16]: f,ax=plt.subplots(figsize=(6,8))
ax=sns.countplot(x='target',data=df)
```



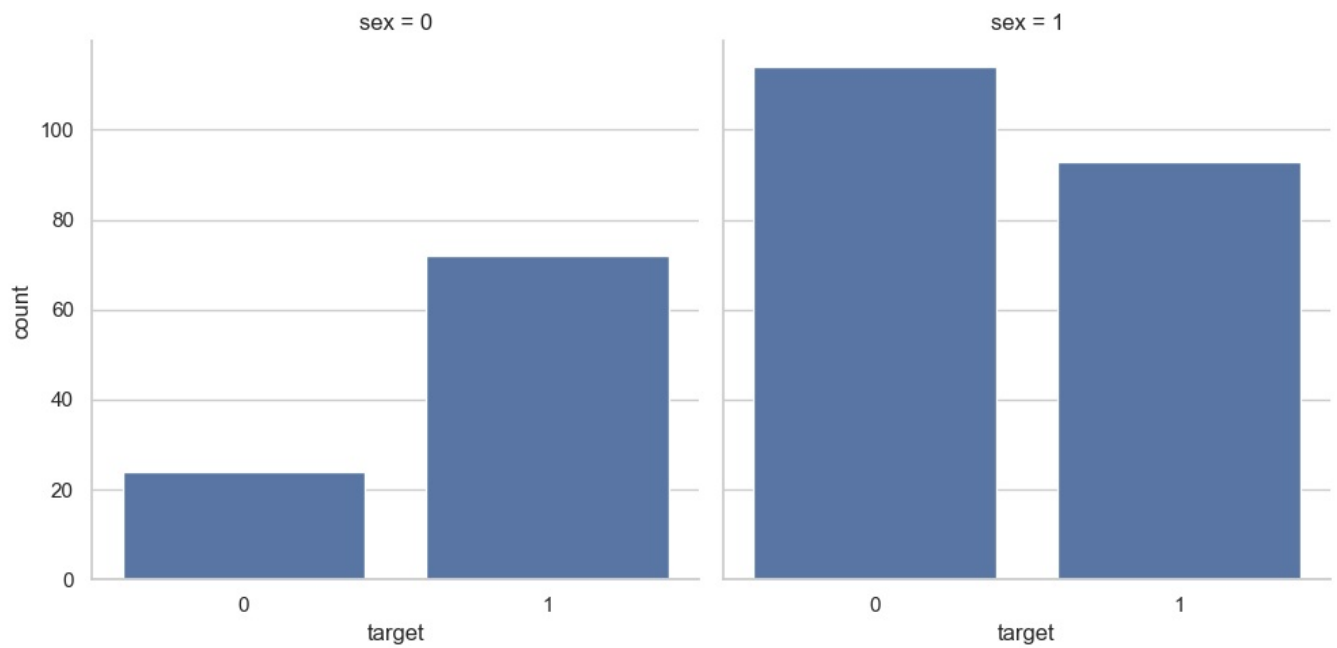
```
In [17]: df.groupby('sex')['target'].value_counts()
```

```
Out[17]: sex  target
0      1         72
      0         24
1      0        114
      1         93
Name: count, dtype: int64
```

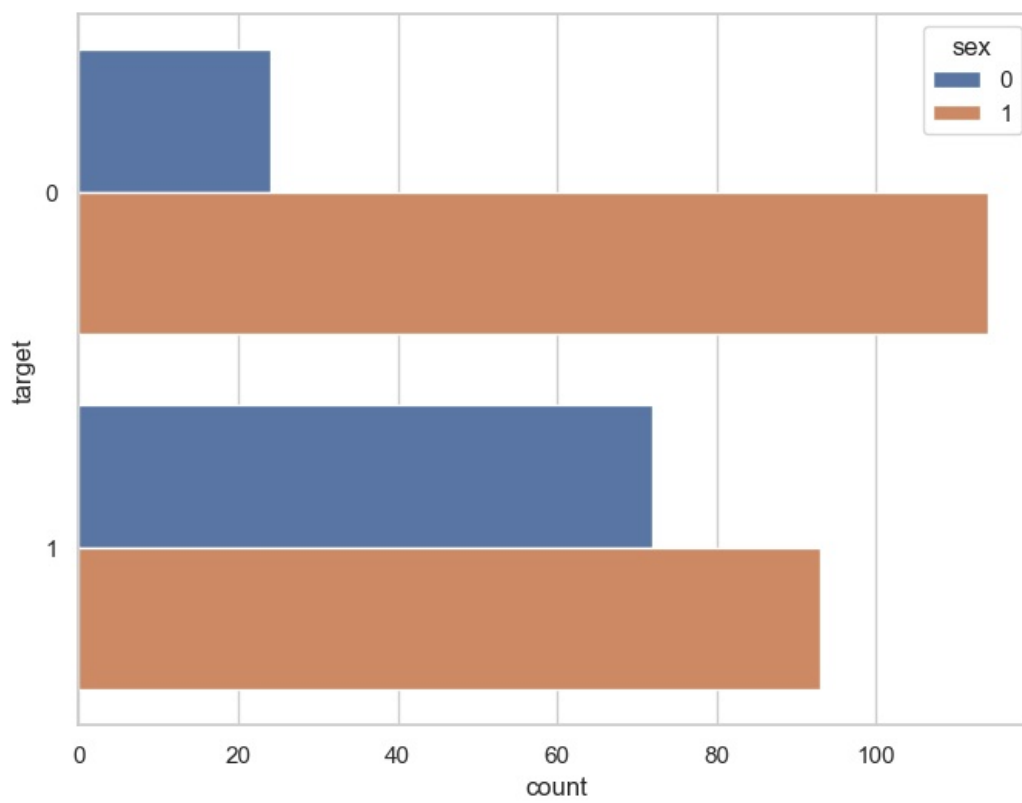
```
In [18]: f,ax=plt.subplots(figsize=(8,6))
ax=sns.countplot(x='sex',hue='target',data=df)
sns.set(style='whitegrid')
```



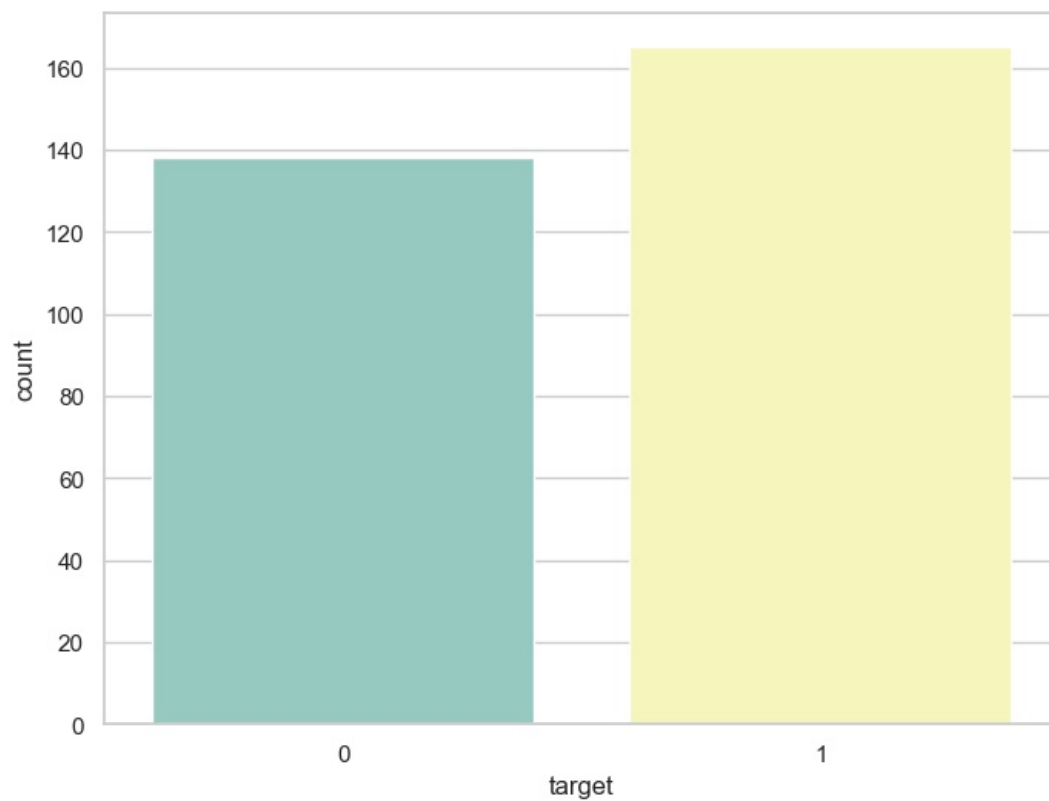
```
In [19]: ax=sns.catplot(x='target',col='sex',data=df,kind='count',height=5,aspect=1)
```



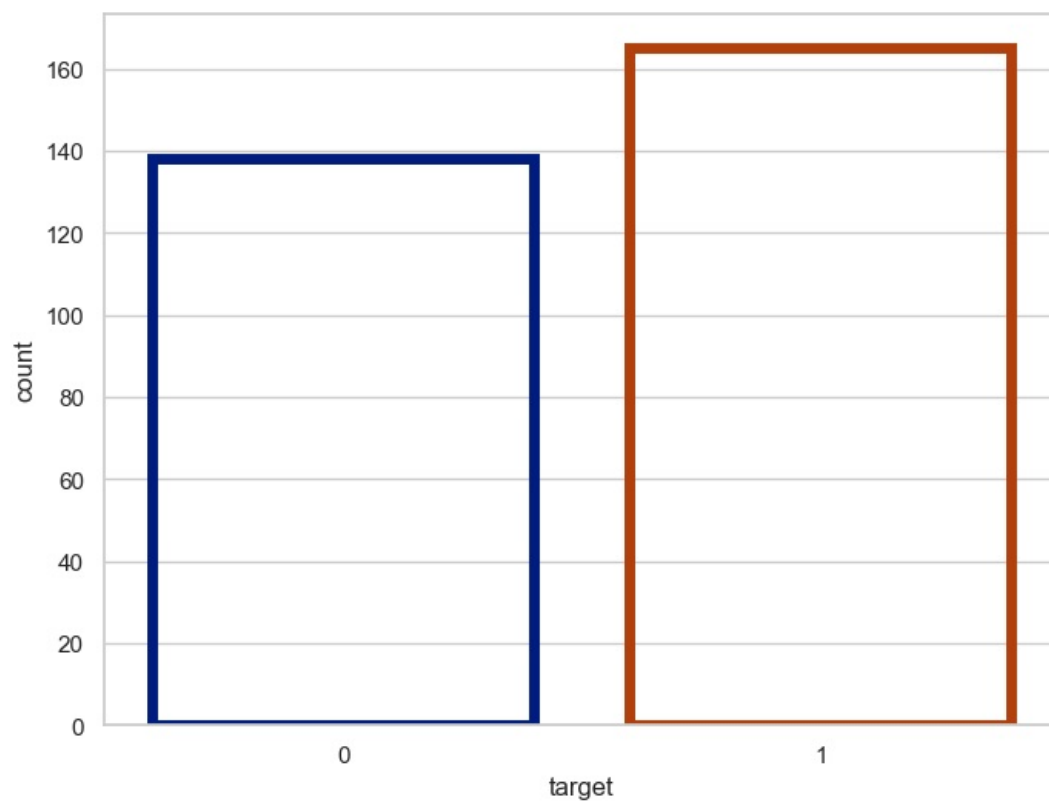
```
In [20]: f,ax=plt.subplots(figsize=(8,6))
ax=sns.countplot(y='target',hue='sex',data=df)
```



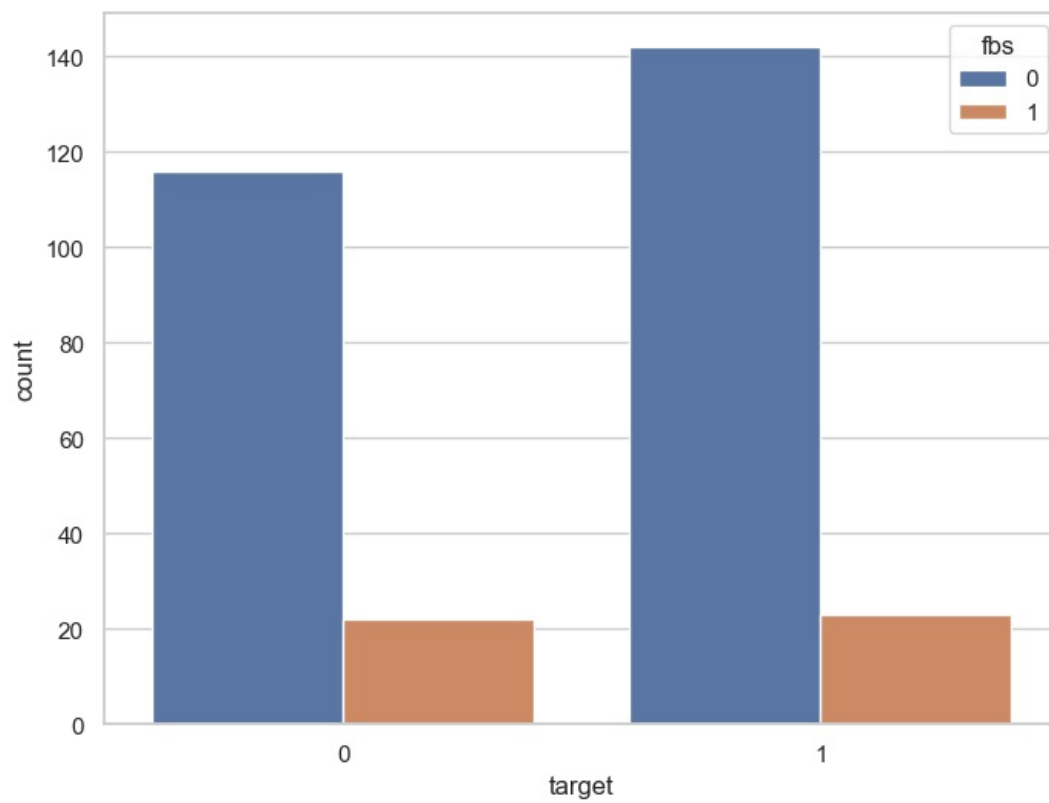
```
In [21]: f,ax=plt.subplots(figsize=(8,6))
ax=sns.countplot(x='target',data=df,palette='Set3')
```



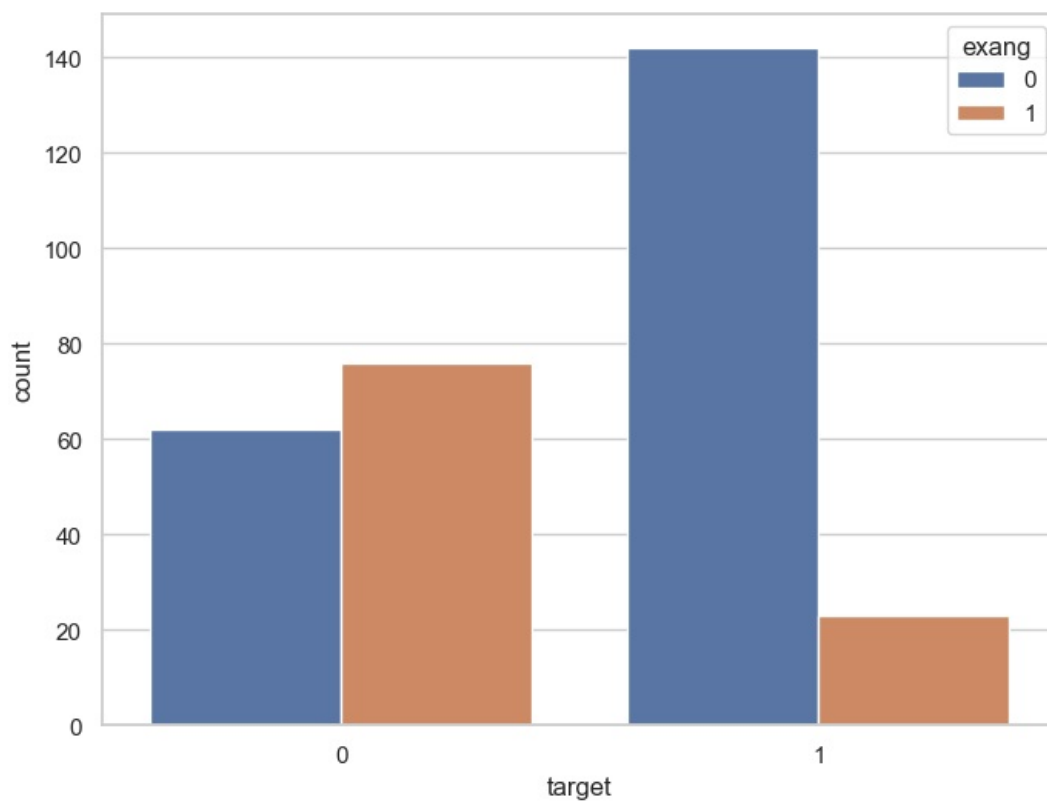
```
In [22]: f,ax=plt.subplots(figsize=(8,6))
ax=sns.countplot(x='target',data=df,facecolor=(0,0,0,0),linewidth=5,edgecolor=sns.color_palette('dark',3))
```



```
In [23]: f,ax=plt.subplots(figsize=(8,6))  
ax=sns.countplot(x='target',hue='fbs',data=df)
```



```
In [24]: f,ax=plt.subplots(figsize=(8,6))  
ax=sns.countplot(x='target',hue='exang',data=df)
```



```
In [25]: correlation=df.corr()
```

```
In [26]: correlation['target'].sort_values(ascending=False)
```

```
Out[26]: target      1.000000
cp          0.433798
thalach     0.421741
slope       0.345877
restecg     0.137230
fbs         -0.028046
chol        -0.085239
trestbps    -0.144931
age         -0.225439
sex         -0.280937
thal        -0.344029
ca          -0.391724
oldpeak     -0.430696
exang       -0.436757
Name: target, dtype: float64
```

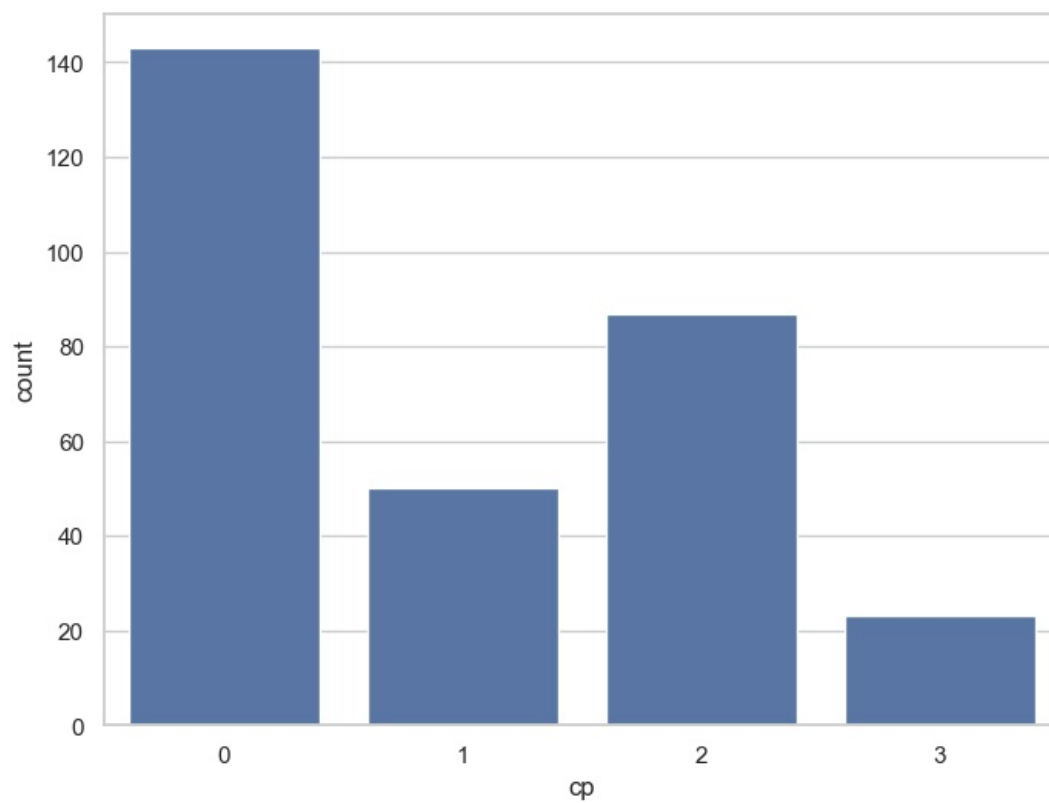
```
In [27]: correlation['target'].sort_values(ascending=True)
```

```
Out[27]: exang       -0.436757
oldpeak    -0.430696
ca         -0.391724
thal       -0.344029
sex        -0.280937
age        -0.225439
trestbps   -0.144931
chol       -0.085239
fbs        -0.028046
restecg    0.137230
slope      0.345877
thalach    0.421741
cp         0.433798
target     1.000000
Name: target, dtype: float64
```

```
In [28]: df['cp'].value_counts()
```

```
Out[28]: cp
0      143
2       87
1       50
3       23
Name: count, dtype: int64
```

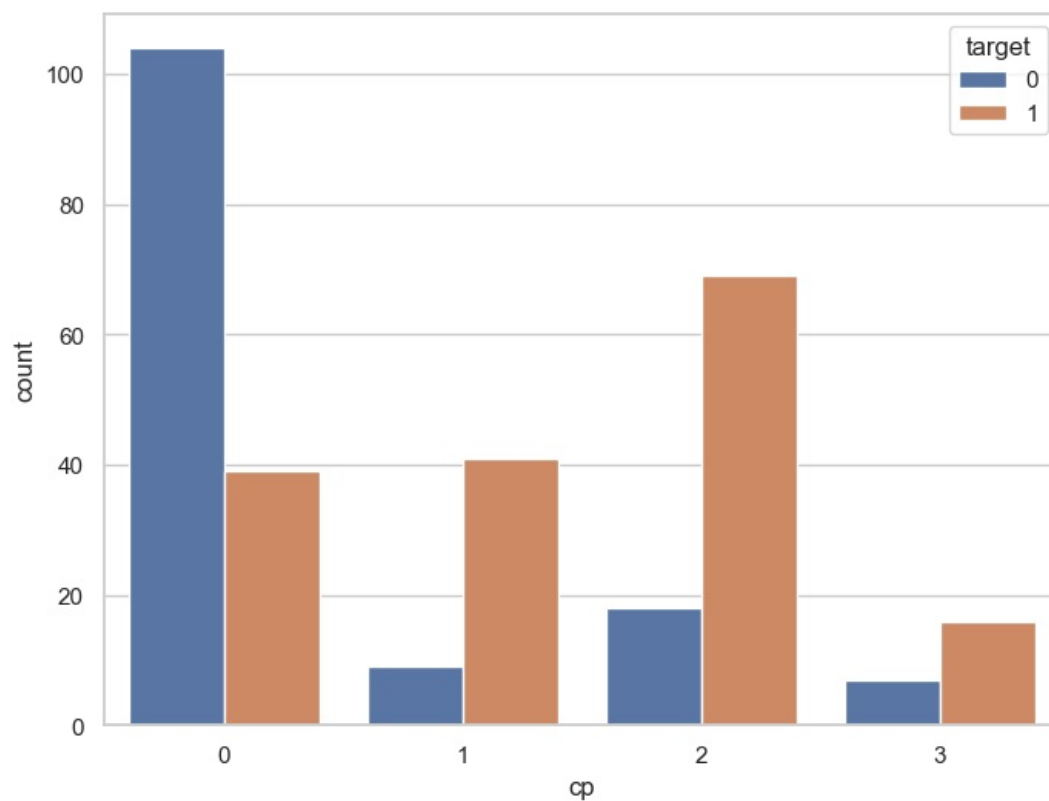
```
In [29]: f,ax=plt.subplots(figsize=(8,6))
ax=sns.countplot(x='cp',data=df)
```



```
In [30]: df.groupby('cp')['target'].value_counts()
```

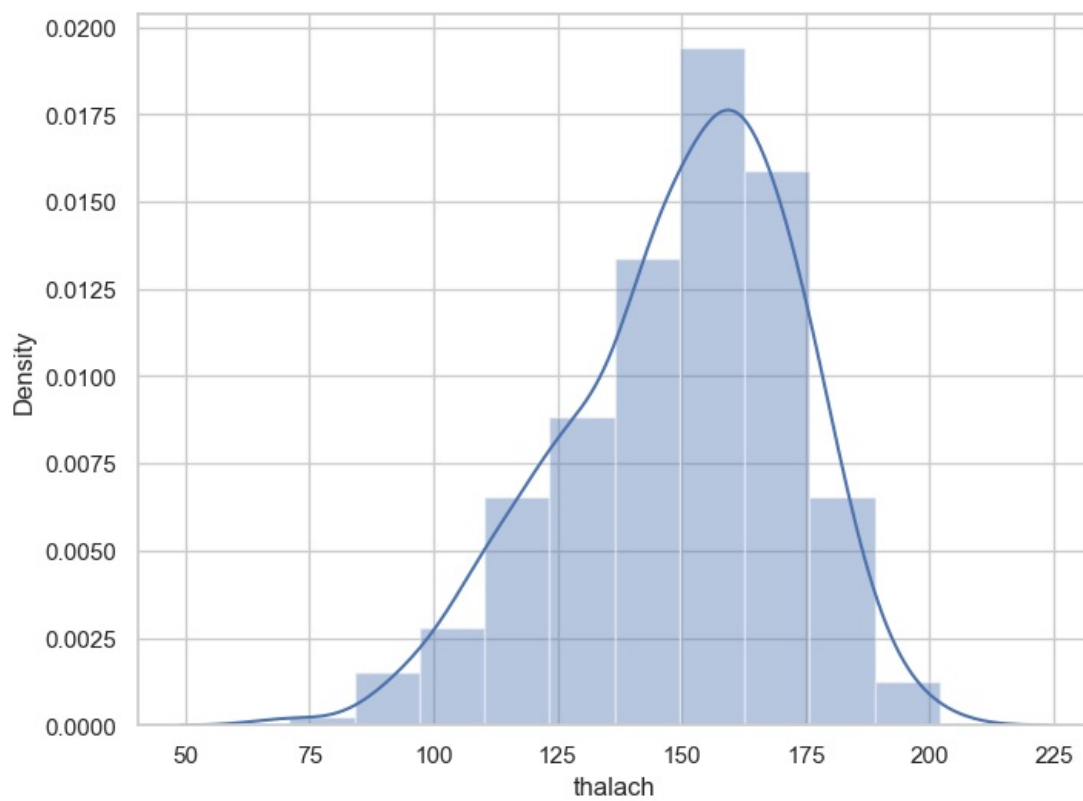
```
Out[30]: cp target
0  0      104
   1       39
1  1       41
   0        9
2  1       69
   0       18
3  1       16
   0        7
Name: count, dtype: int64
```

```
In [31]: f,ax=plt.subplots(figsize=(8,6))
ax=sns.countplot(x='cp',hue='target',data=df)
```

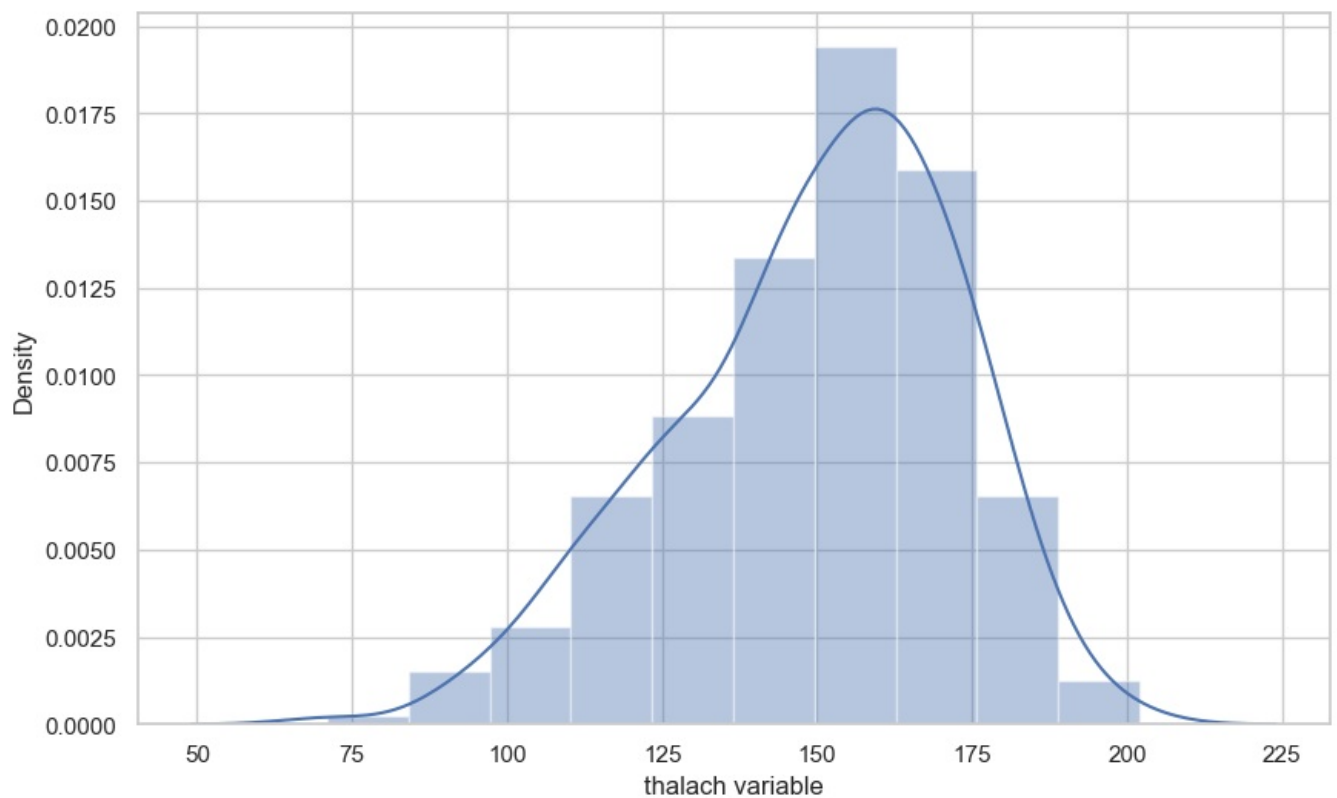


```
In [32]: f,ax=plt.subplots(figsize=(8,6))
x=df['thalach']
ax=sns.distplot(x,bins=10)
```

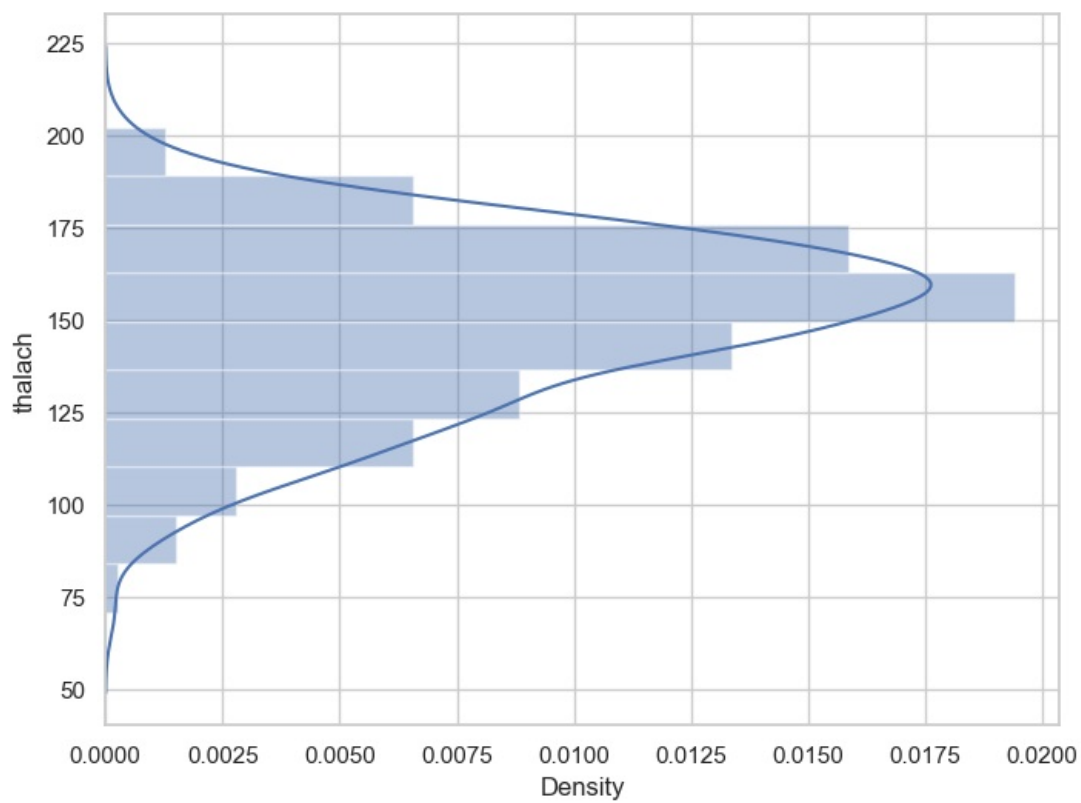




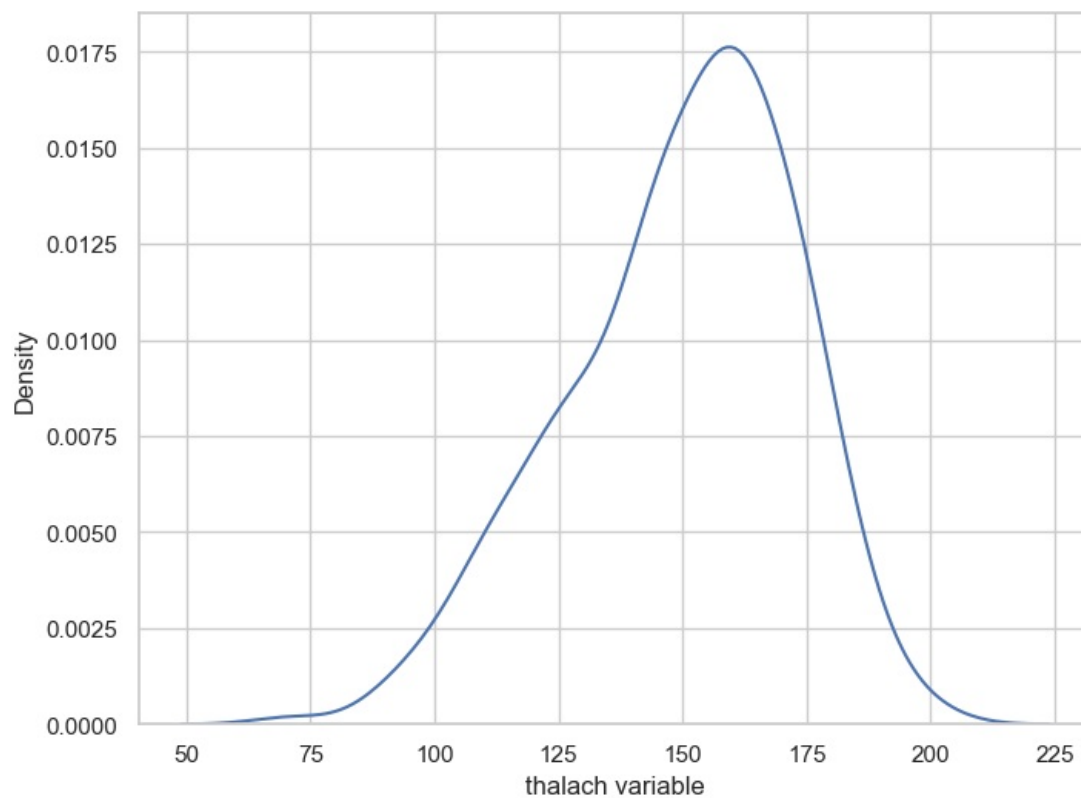
```
In [33]: f,ax=plt.subplots(figsize=(10,6))
x=df['thalach']
x=pd.Series(x,name='thalach variable')
ax=sns.distplot(x,bins=10)
```



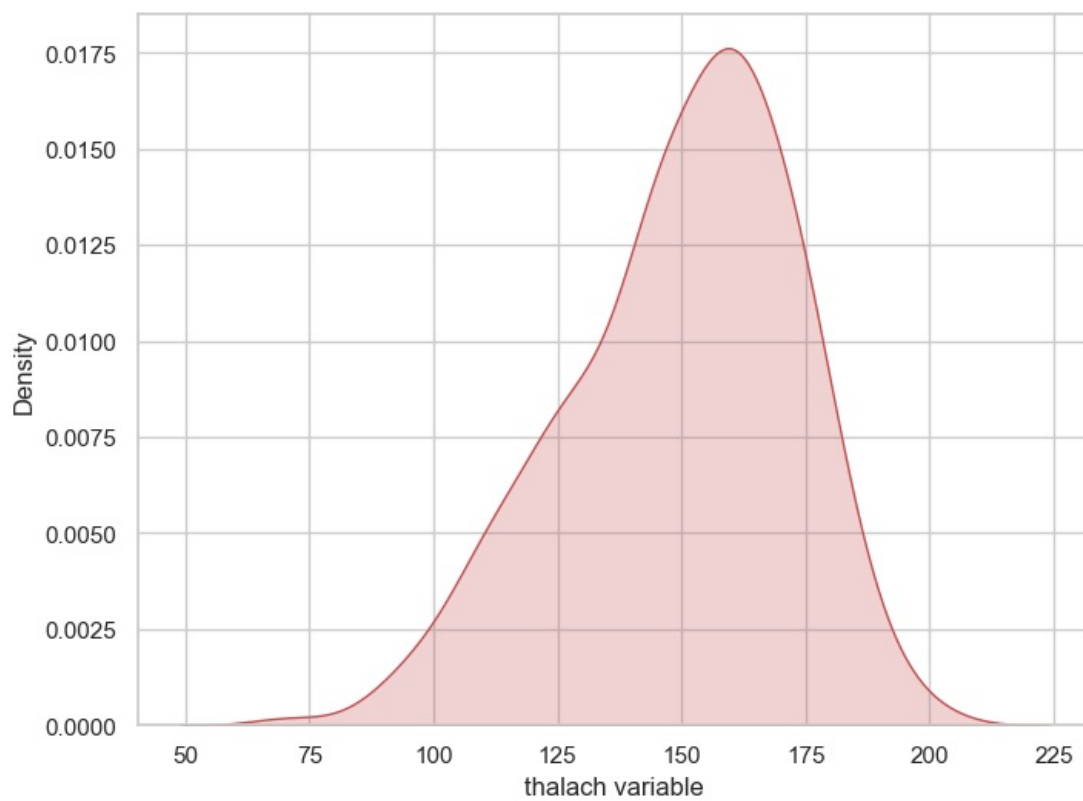
```
In [34]: f,ax=plt.subplots(figsize=(8,6))
x=df['thalach']
ax=sns.distplot(x,bins=10,vertical=True)
```



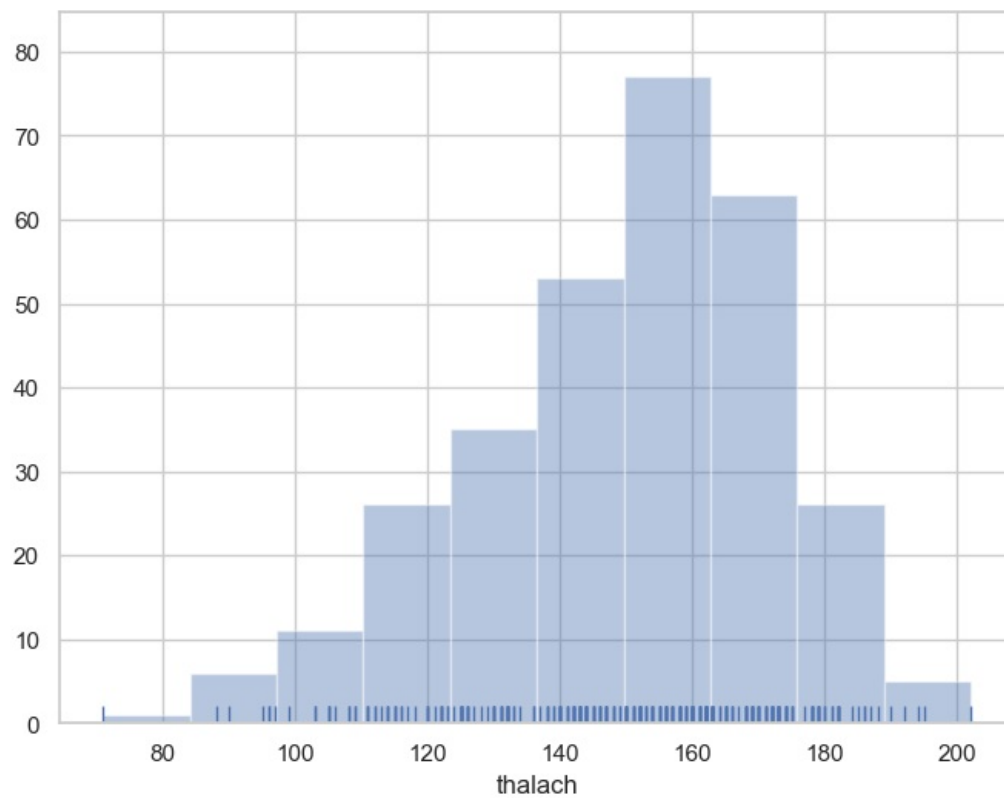
```
In [35]: f,ax=plt.subplots(figsize=(8,6))
x=df['thalach']
x=pd.Series(x,name='thalach variable')
ax=sns.kdeplot(x)
```



```
In [36]: f,ax=plt.subplots(figsize=(8,6))
x=df['thalach']
x=pd.Series(x,name='thalach variable')
ax=sns.kdeplot(x,shade=True,color='r')
```

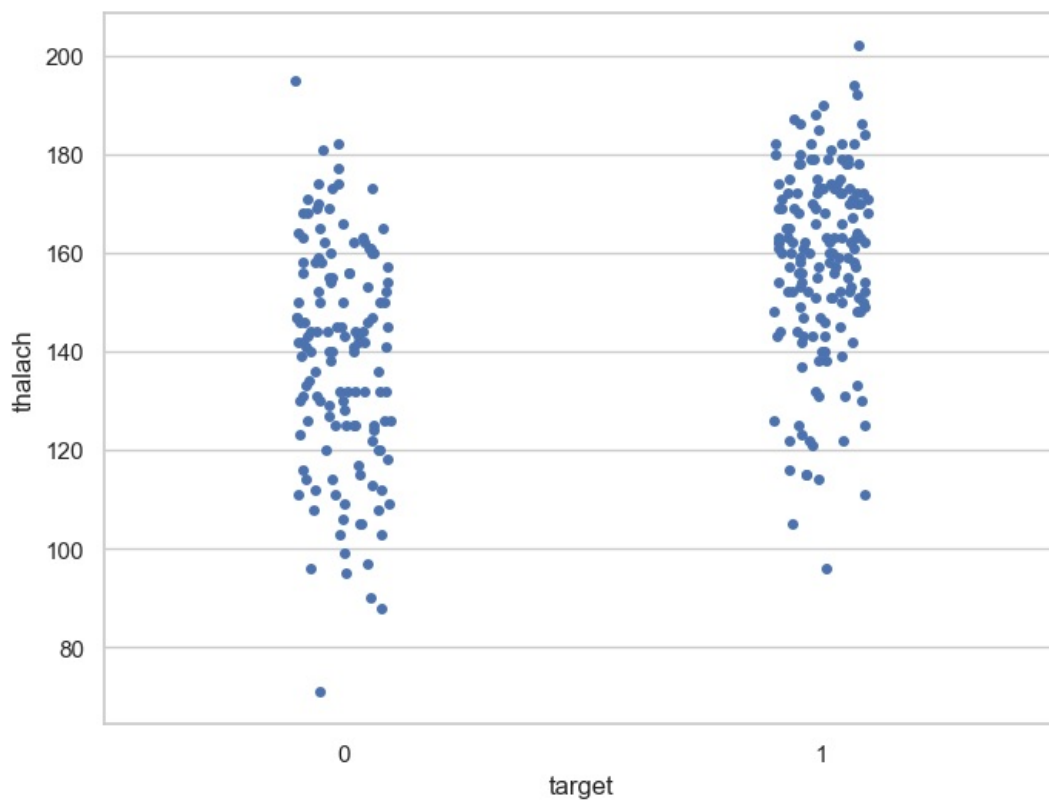


```
In [37]: f,ax=plt.subplots(figsize=(8,6))
x=df['thalach']
ax=sns.distplot(x,kde=False,rug=True,bins=10)
```



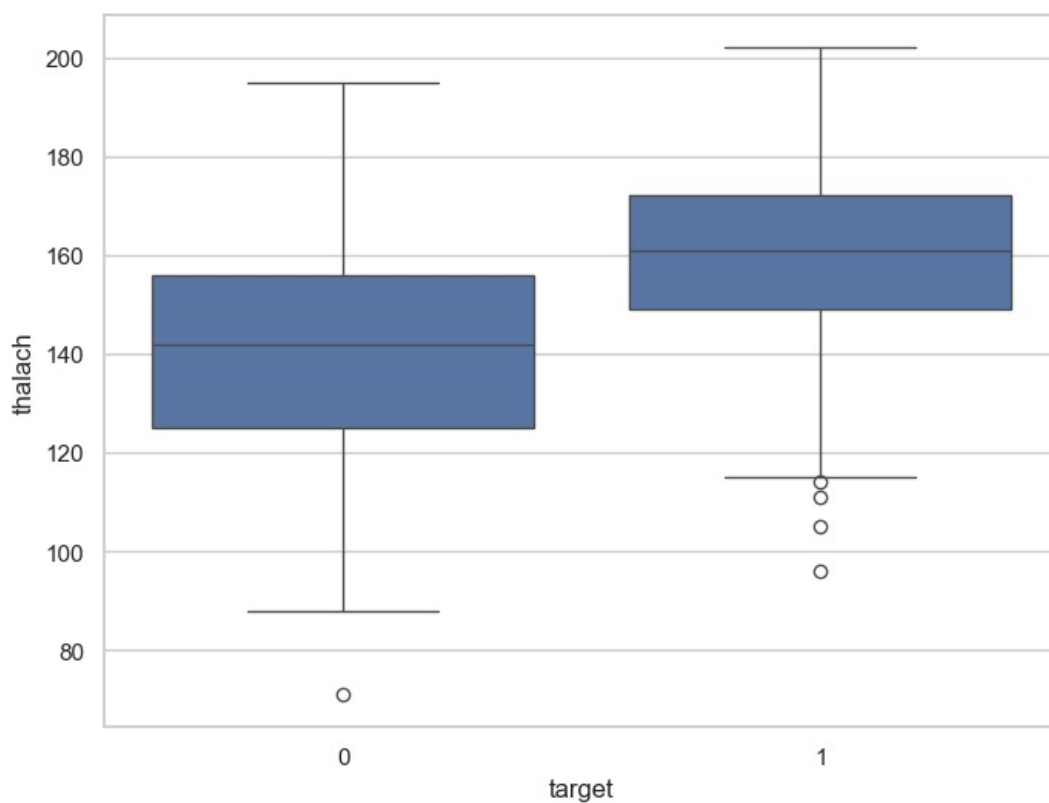
```
In [38]: f,ax=plt.subplots(figsize=(8,6))
sns.stripplot(x='target',y='thalach',data=df)
```

```
Out[38]: <Axes: xlabel='target', ylabel='thalach'>
```



```
In [39]: f,ax=plt.subplots(figsize=(8,6))
sns.boxplot(x='target',y='thalach',data=df)
```

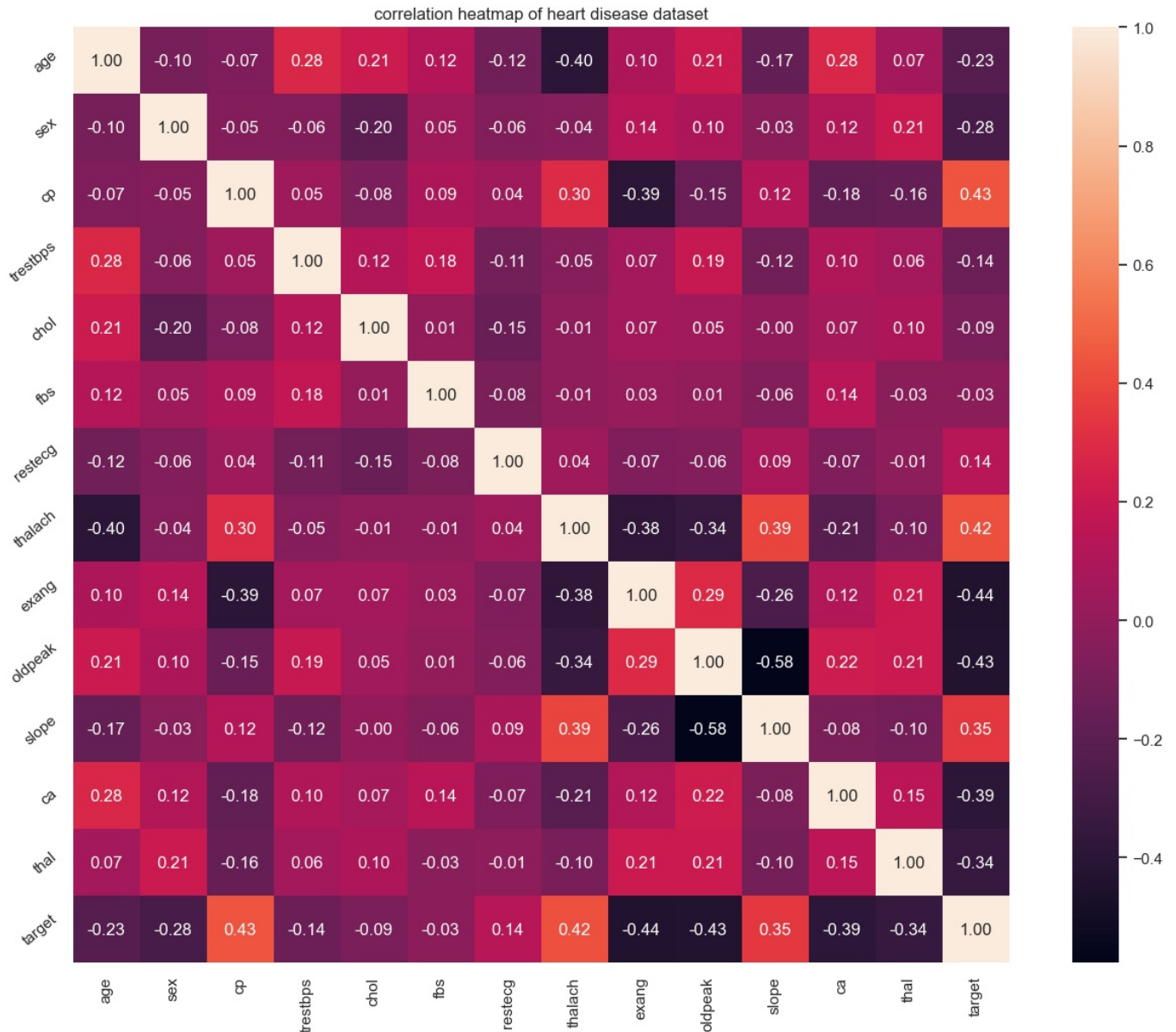
```
Out[39]: <Axes: xlabel='target', ylabel='thalach'>
```



```
In [40]: plt.figure(figsize=(16,12))
plt.title('correlation heatmap of heart disease dataset')
a=sns.heatmap(correlation,square=True,annot=True,fmt='.2f',linecolor='white')
```

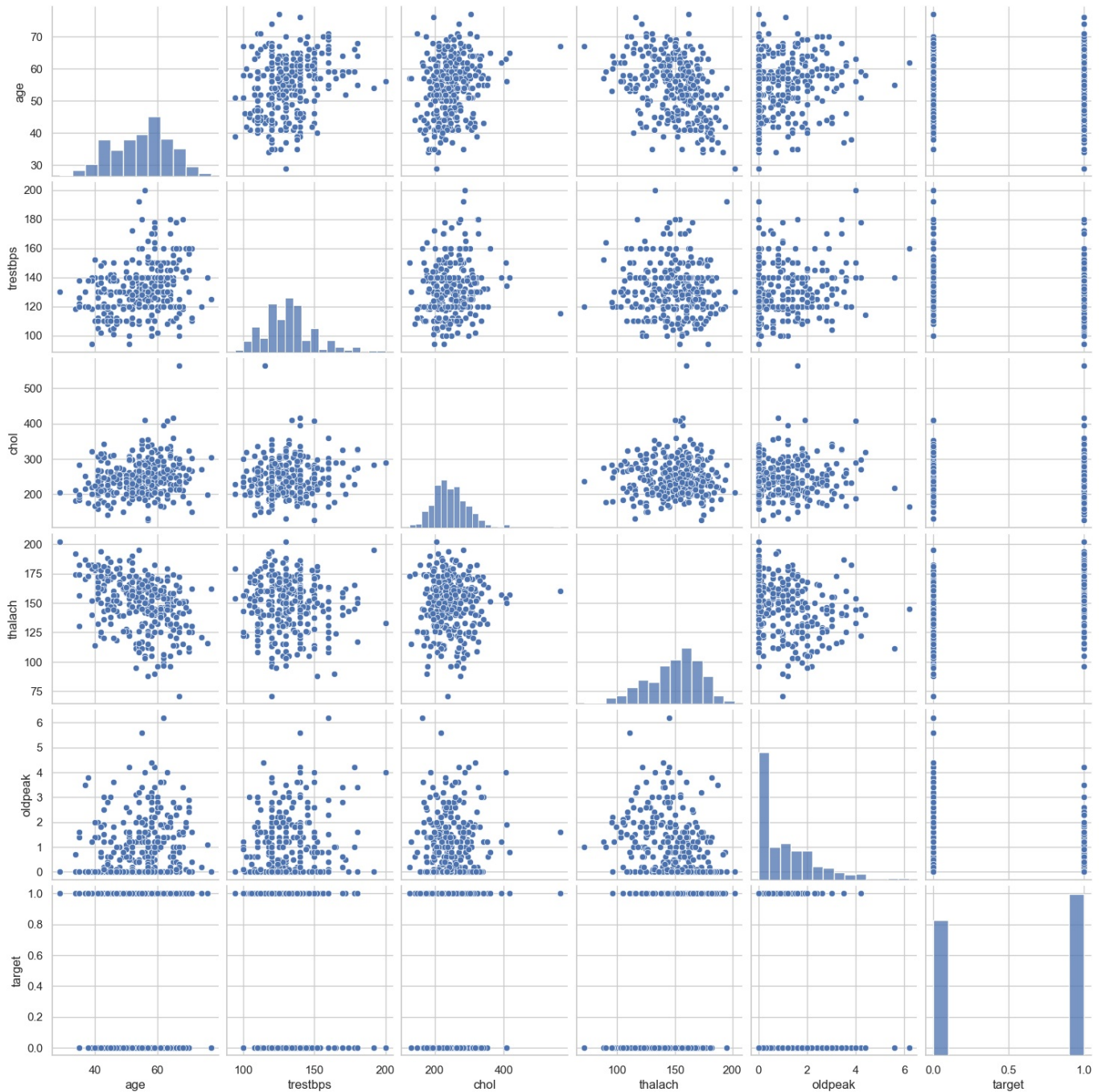
```
a.set_xticklabels(a.get_xticklabels(),rotation=90)
a.set_yticklabels(a.get_yticklabels(),rotation=40)
```

```
Out[40]: [Text(0, 0.5, 'age'),
Text(0, 1.5, 'sex'),
Text(0, 2.5, 'cp'),
Text(0, 3.5, 'trestbps'),
Text(0, 4.5, 'chol'),
Text(0, 5.5, 'fbs'),
Text(0, 6.5, 'restecg'),
Text(0, 7.5, 'thalach'),
Text(0, 8.5, 'exang'),
Text(0, 9.5, 'oldpeak'),
Text(0, 10.5, 'slope'),
Text(0, 11.5, 'ca'),
Text(0, 12.5, 'thal'),
Text(0, 13.5, 'target')]
```



```
In [41]: num_var=['age','trestbps','chol','thalach','oldpeak','target']
sns.pairplot(df[num_var],kind='scatter',diag_kind='hist')
```

```
Out[41]: <seaborn.axisgrid.PairGrid at 0x26580176b40>
```



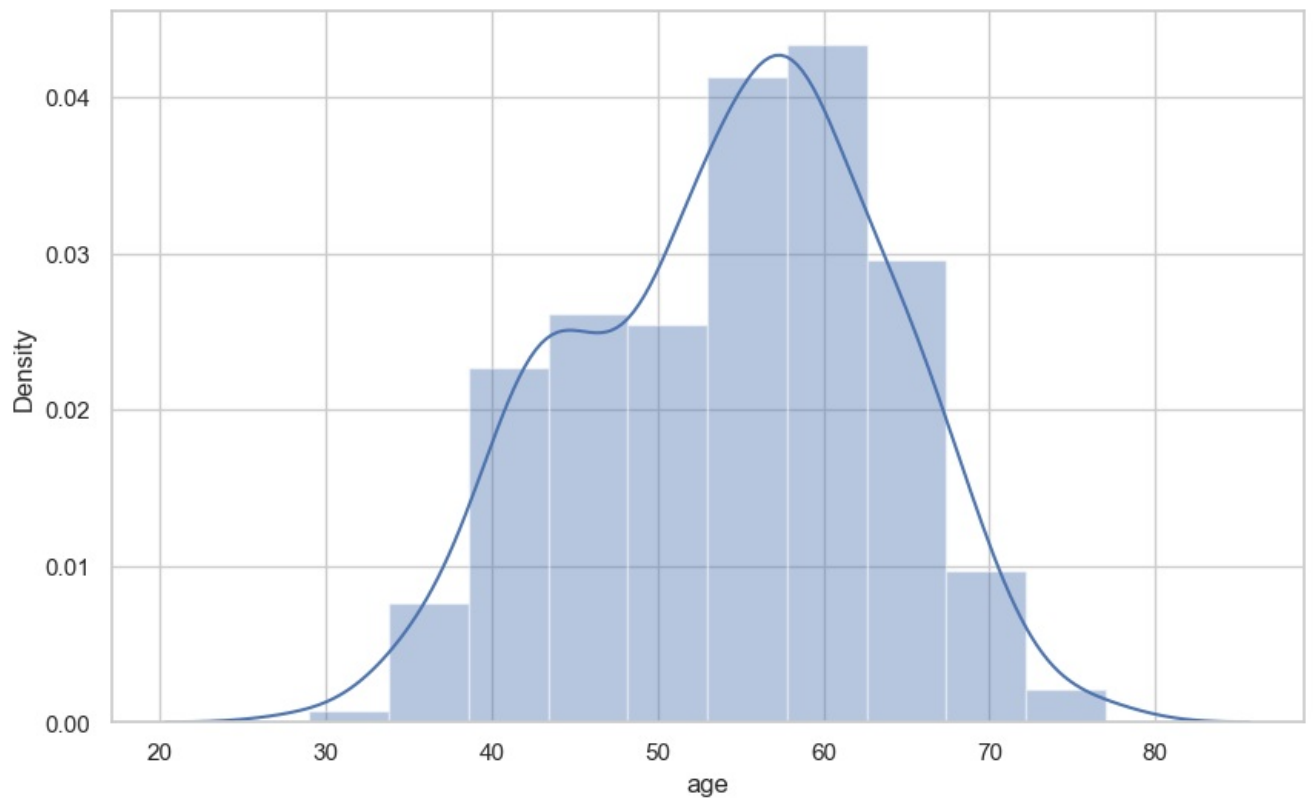
```
In [42]: df['age'].nunique()
```

```
Out[42]: 41
```

```
In [43]: df['age'].describe()
```

```
Out[43]: count    303.000000
         mean     54.366337
         std      9.082101
         min     29.000000
         25%     47.500000
         50%     55.000000
         75%     61.000000
         max     77.000000
         Name: age, dtype: float64
```

```
In [44]: f,ax=plt.subplots(figsize=(10,6))
x=df['age']
ax=sns.distplot(x,bins=10)
```



In [ ]:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js