

In [1]:

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
sns.set(style="whitegrid")
```

In [2]:

```
heart=pd.read_csv(r'C:\Users\Dell\Downloads\15th\15th\EDA\heart.csv')
```

In [3]:

```
heart
```

Out[3]:

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

303 rows × 14 columns

In [4]:

```
heart.shape
```

Out[4]:

(303, 14)

In [5]:

```
heart.head()
```

Out[5]:

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

In [6]:

```
heart.tail()
```

Out[6]:

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

In [7]:

heart[20:100]

Out[7]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	tar
20	59	1	0	135	234	0	1	161	0	0.5	1	0	3	
21	44	1	2	130	233	0	1	179	1	0.4	2	0	2	
22	42	1	0	140	226	0	1	178	0	0.0	2	0	2	
23	61	1	2	150	243	1	1	137	1	1.0	1	0	2	
24	40	1	3	140	199	0	1	178	1	1.4	2	0	3	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
95	53	1	0	142	226	0	0	111	1	0.0	2	0	3	
96	62	0	0	140	394	0	0	157	0	1.2	1	0	2	
97	52	1	0	108	233	1	1	147	0	0.1	2	3	3	
98	43	1	2	130	315	0	1	162	0	1.9	2	1	2	
99	53	1	2	130	246	1	0	173	0	0.0	2	3	2	

80 rows × 14 columns

In [8]:

heart.describe()

Out[8]:

	age	sex	cp	trestbps	chol	fbs	restecg
count	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
std	9.082101	0.466011	1.032052	17.538143	51.830751	0.356198	0.525860
min	29.000000	0.000000	0.000000	94.000000	126.000000	0.000000	0.000000
25%	47.500000	0.000000	0.000000	120.000000	211.000000	0.000000	0.000000
50%	55.000000	1.000000	1.000000	130.000000	240.000000	0.000000	1.000000
75%	61.000000	1.000000	2.000000	140.000000	274.500000	0.000000	1.000000
max	77.000000	1.000000	3.000000	200.000000	564.000000	1.000000	2.000000

In [9]:

```
heart.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]:

```
heart.columns
```

Out[10]:

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

In [12]:

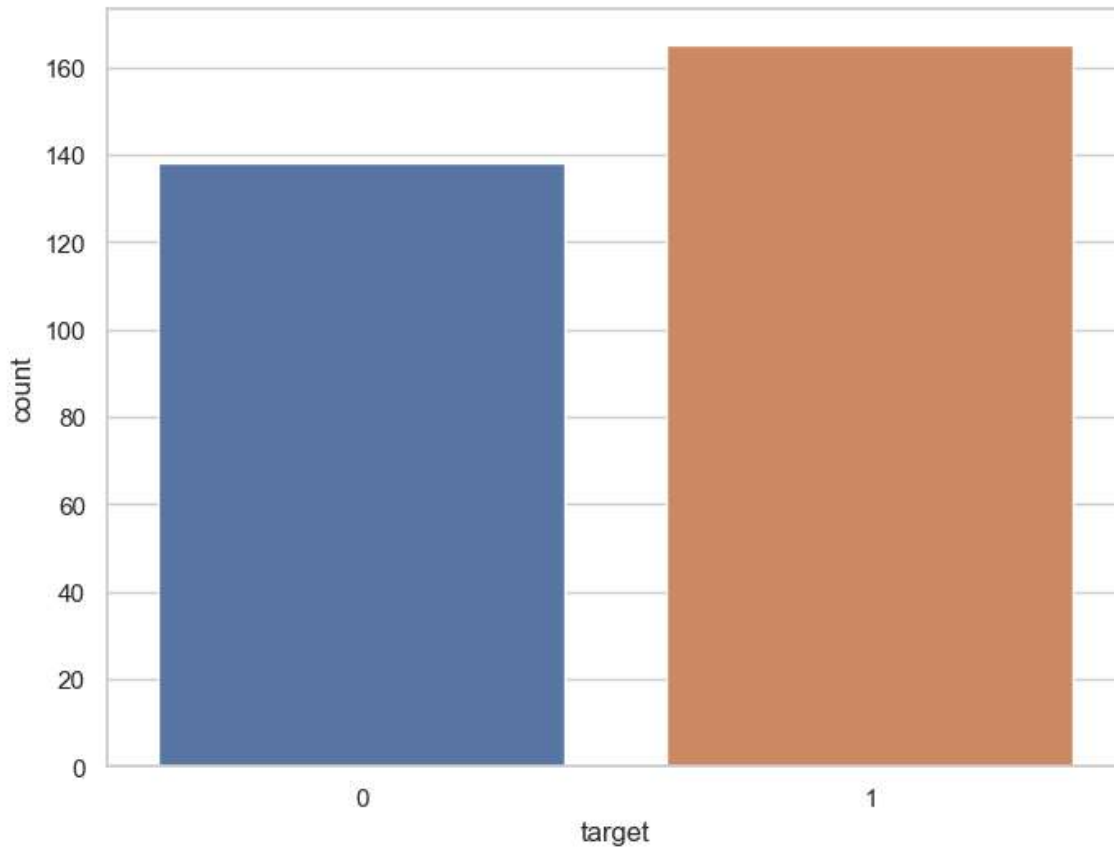
```
heart['target'].value_counts()
```

Out[12]:

```
1    165
0    138
Name: target, dtype: int64
```

In [13]:

```
f, ax = plt.subplots(figsize=(8, 6))  
ax = sns.countplot(x="target", data=heart)  
plt.show()
```



In [15]:

```
heart.groupby('sex')['target'].value_counts()
```

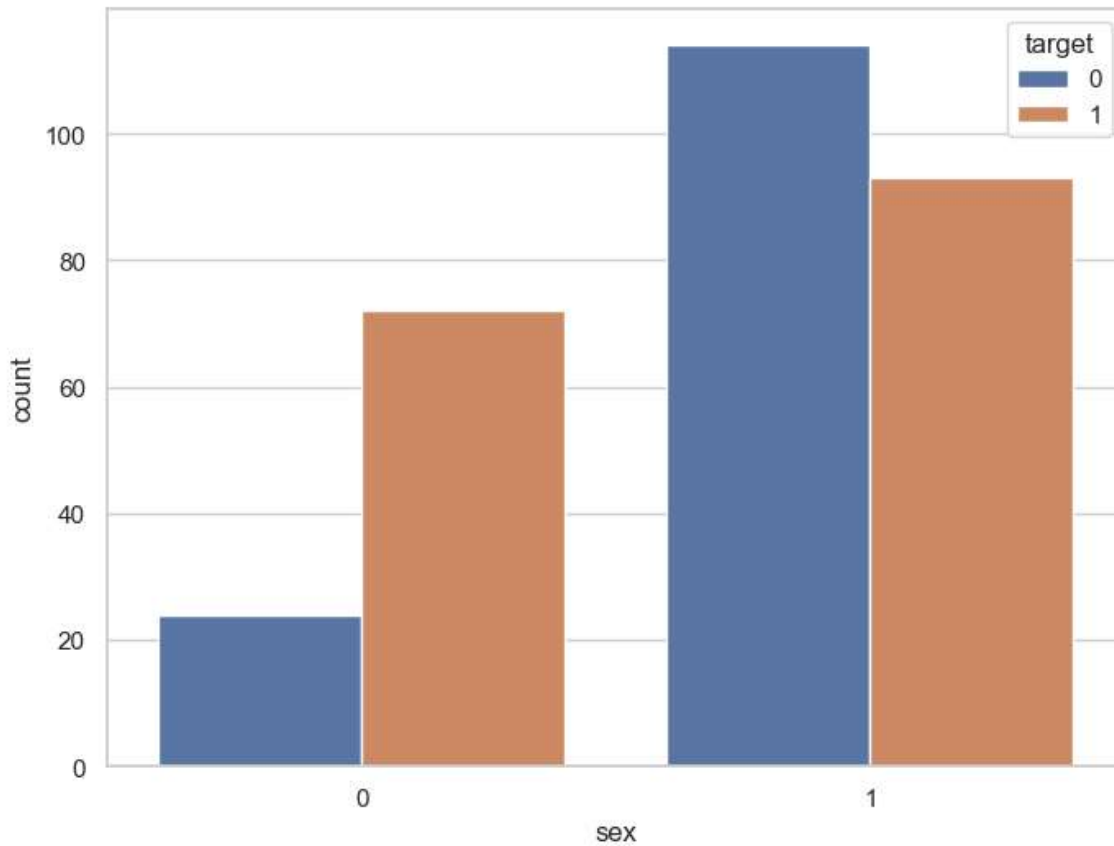
Out[15]:

sex	target	
0	1	72
	0	24
1	0	114
	1	93

Name: target, dtype: int64

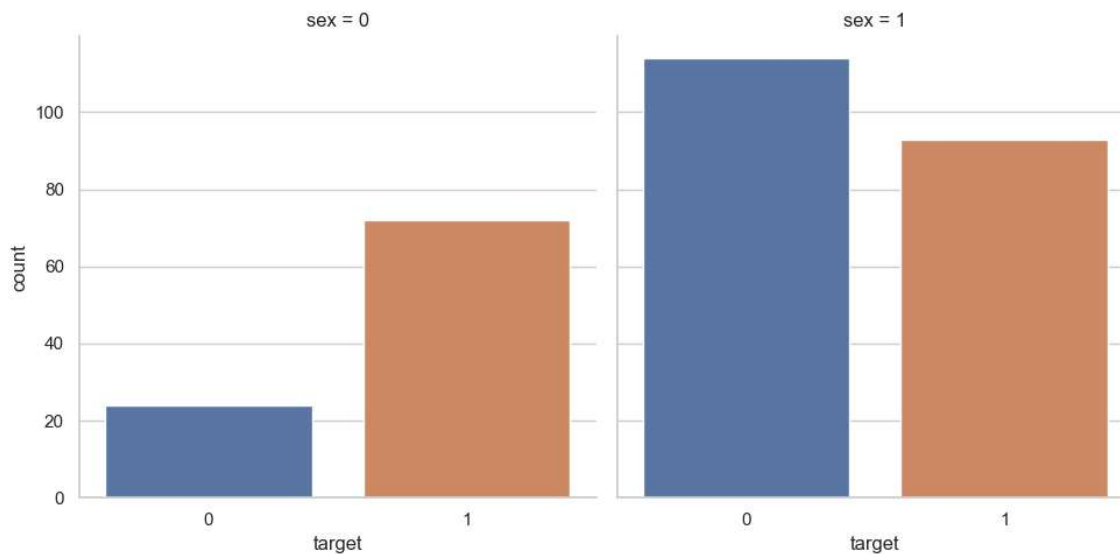
In [16]:

```
f, ax = plt.subplots(figsize=(8, 6))  
ax = sns.countplot(x="sex", hue="target", data=heart)  
plt.show()
```



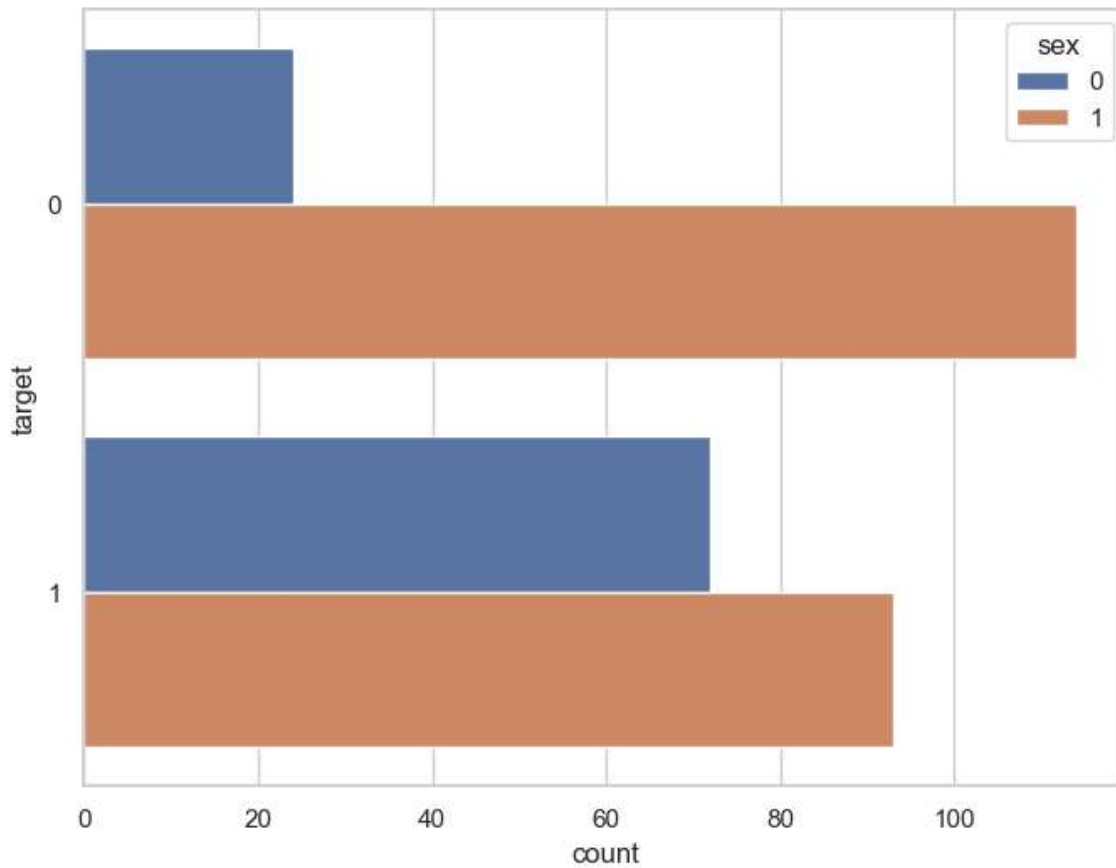
In [18]:

```
vis = sns.catplot(x="target", col="sex", data=heart, kind="count", height=5, aspect=1)
```



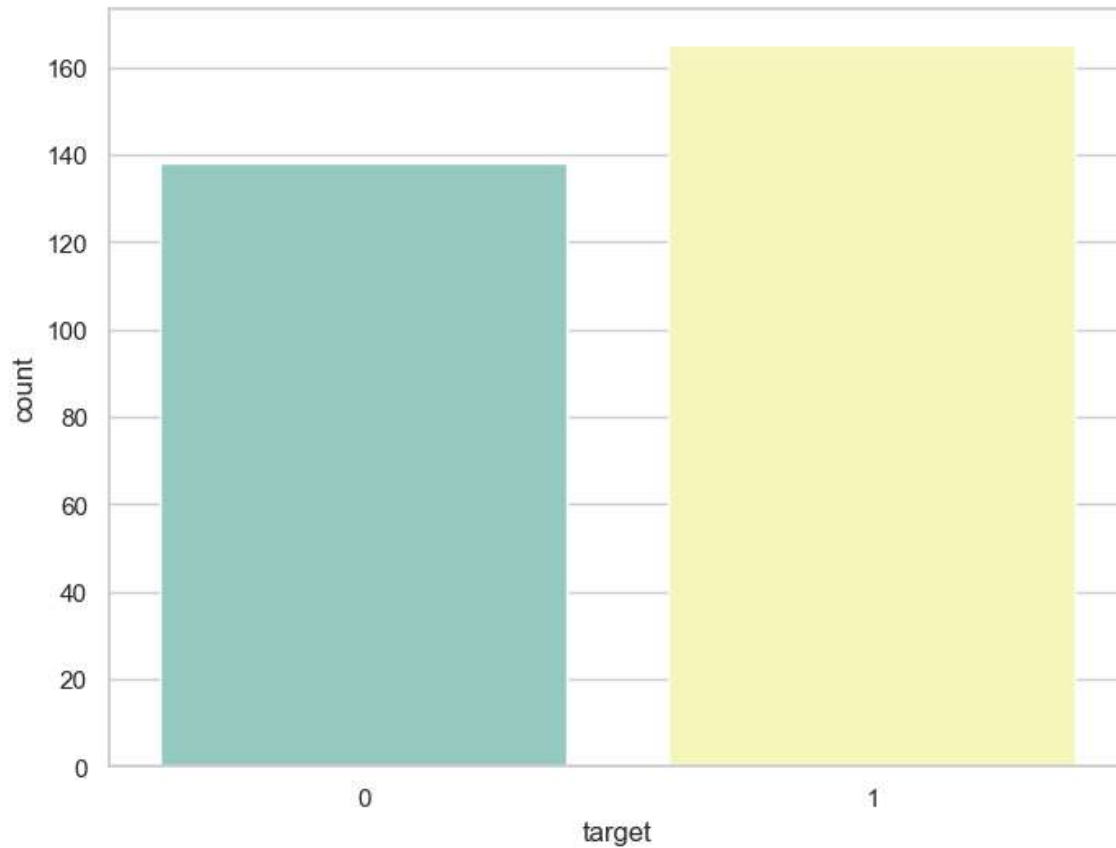
In [19]:

```
f, vis = plt.subplots(figsize=(8, 6))  
ax = sns.countplot(y="target", hue="sex", data=heart)  
plt.show()
```



In [20]:

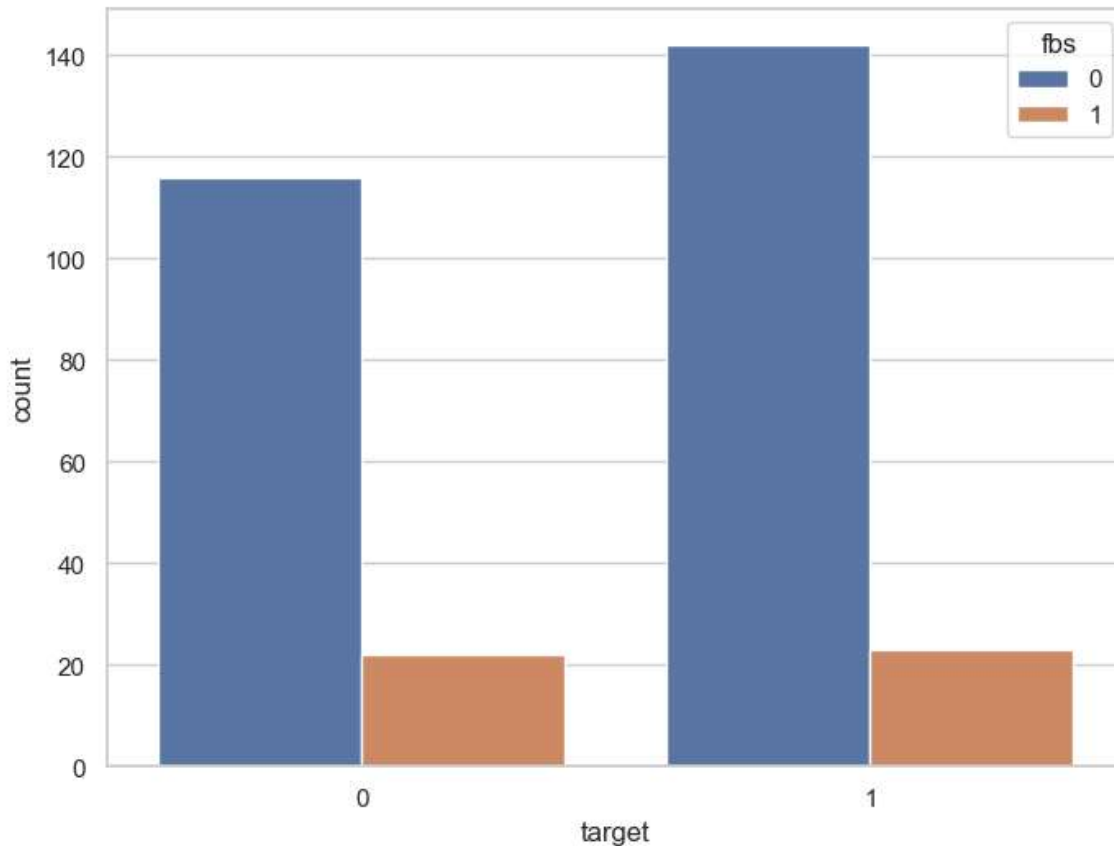
```
f, vis = plt.subplots(figsize=(8, 6))  
ax = sns.countplot(x="target", data=heart, palette="Set3")  
plt.show()
```





In [21]:

```
f, ax = plt.subplots(figsize=(8, 6))  
ax = sns.countplot(x="target", hue="fbs", data=heart)  
plt.show()
```



In [23]:

```
correlation = heart.corr()
```

In [24]:

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

Out[24]:

```
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 [25]:

```
heart['cp'].nunique()
```

Out[25]:

4

In [26]:

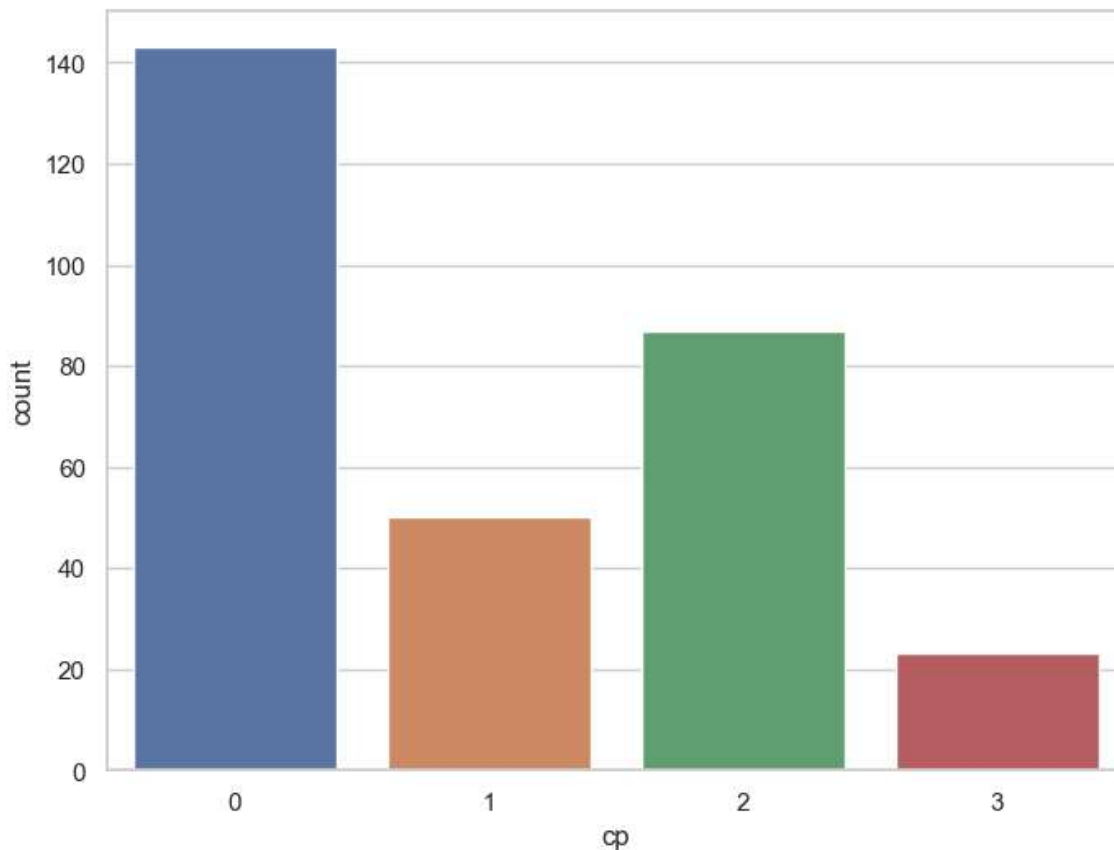
```
heart['cp'].value_counts()
```

Out[26]:

```
0    143
2     87
1     50
3     23
Name: cp, dtype: int64
```

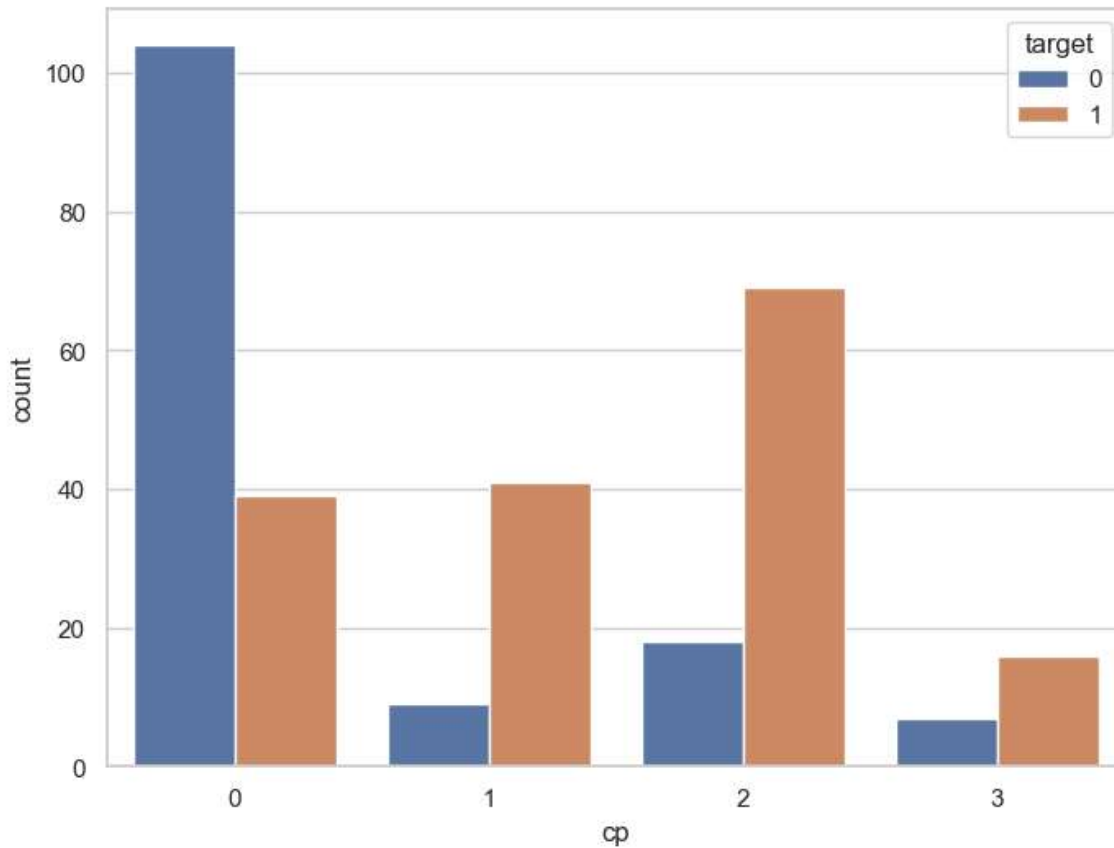
In [27]:

```
f, ax = plt.subplots(figsize=(8, 6))
ax = sns.countplot(x="cp", data=heart)
plt.show()
```



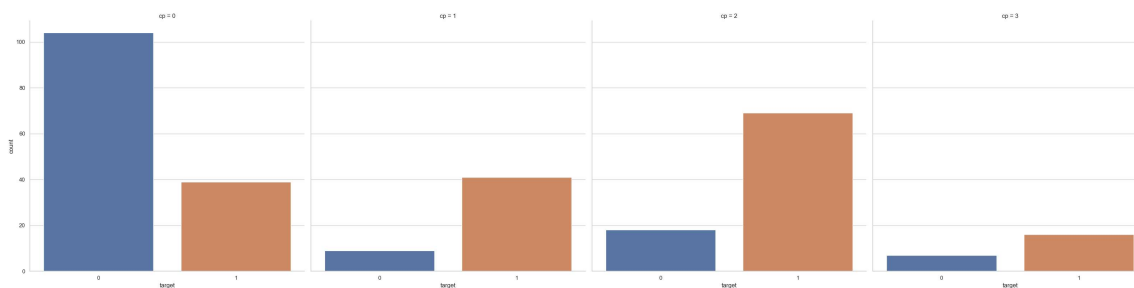
In [28]:

```
f, ax = plt.subplots(figsize=(8, 6))  
ax = sns.countplot(x="cp", hue="target", data=heart)  
plt.show()
```



In [30]:

```
vis= sns.catplot(x="target", col="cp", data=heart, kind="count", height=8, aspect=1)
```



In [31]:

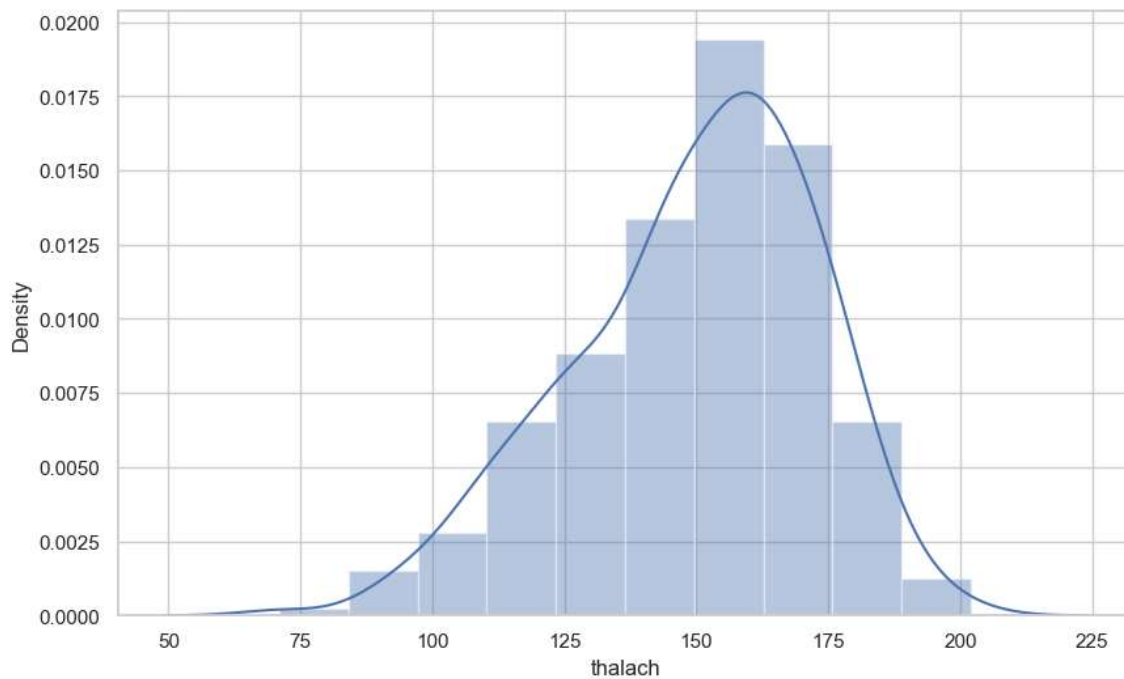
```
heart['thalach'].nunique()
```

Out[31]:

91

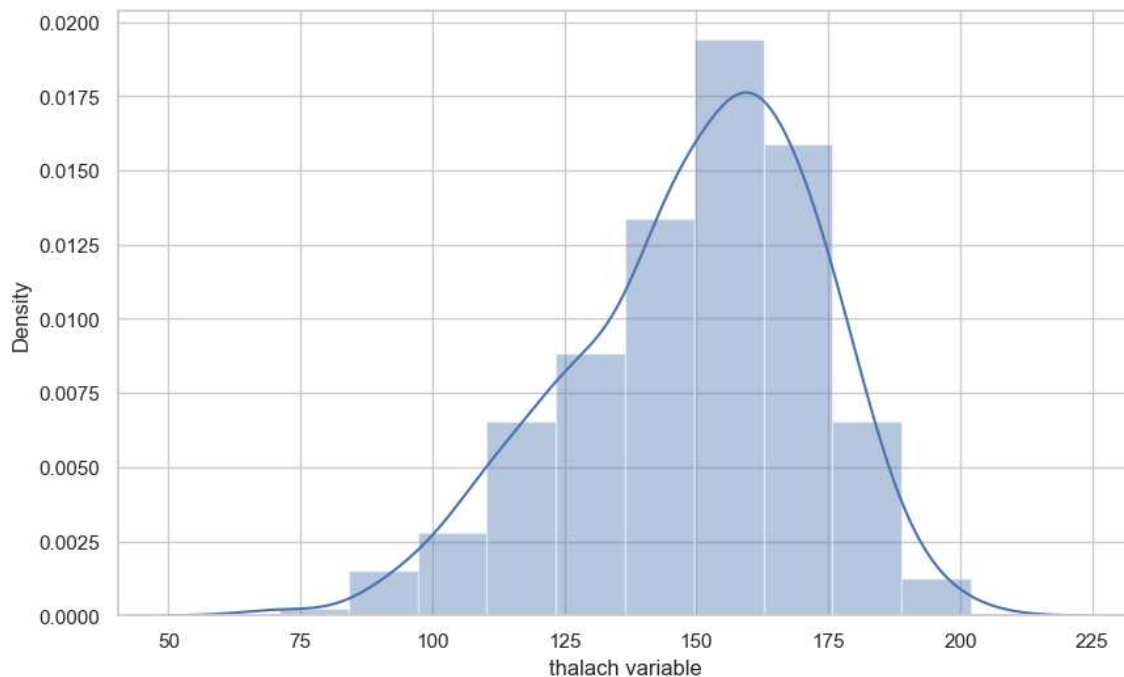
In [32]:

```
f, ax = plt.subplots(figsize=(10,6))
x = heart['thalach']
ax = sns.distplot(x, bins=10)
plt.show()
```



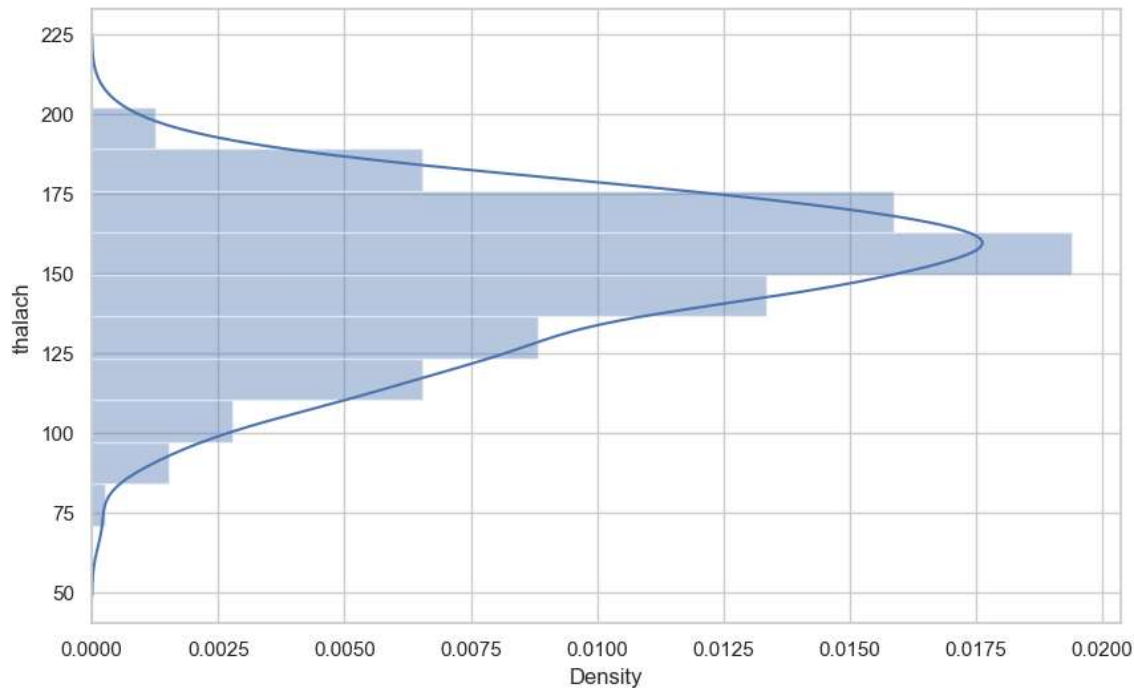
In [33]:

```
f, ax = plt.subplots(figsize=(10,6))
x = heart['thalach']
x = pd.Series(x, name="thalach variable")
ax = sns.distplot(x, bins=10)
plt.show()
```



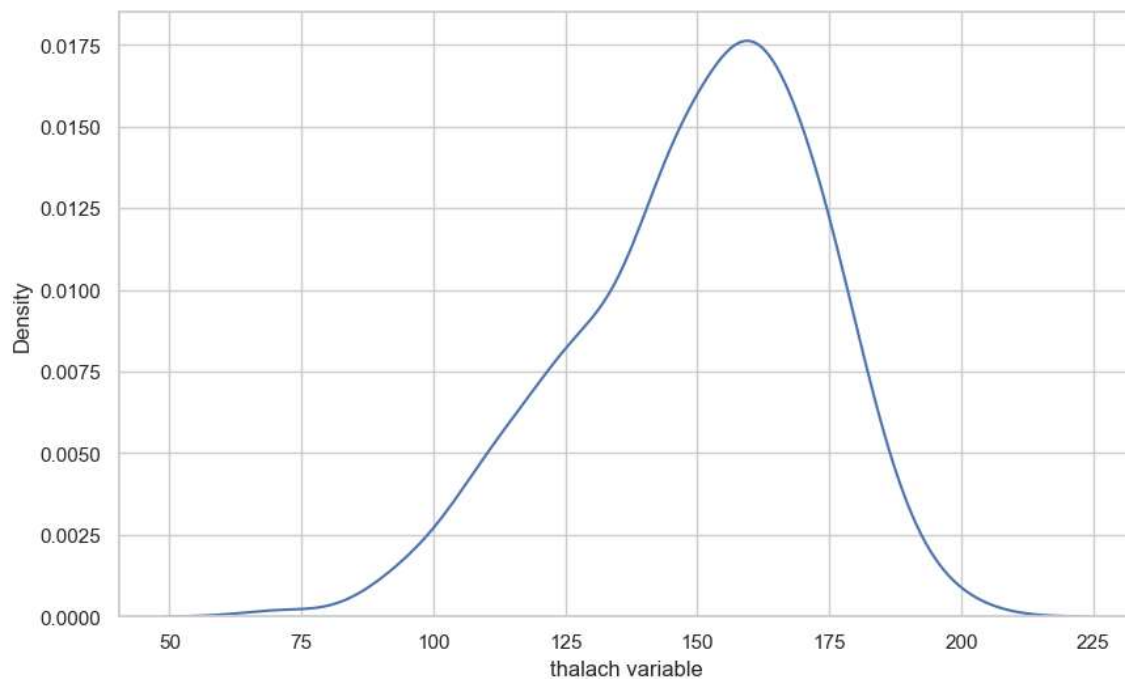
In [34]:

```
f, vis = plt.subplots(figsize=(10,6))
x = heart['thalach']
ax = sns.distplot(x, bins=10, vertical=True)
plt.show()
```



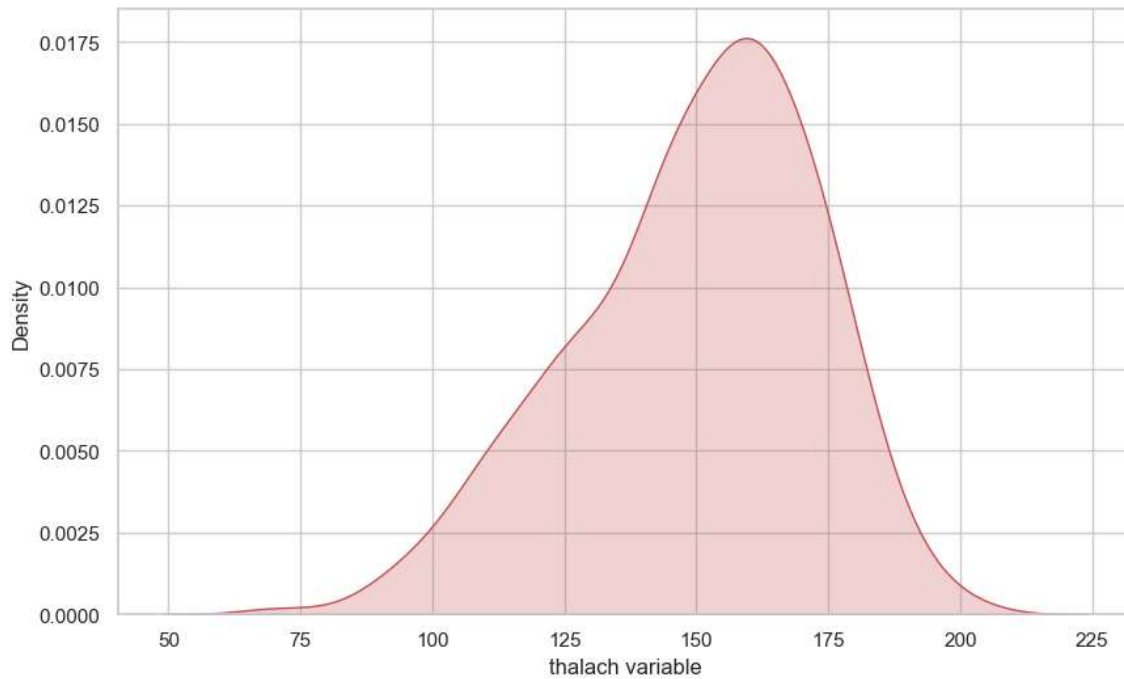
In [35]:

```
f, vis = plt.subplots(figsize=(10,6))
x = heart['thalach']
x = pd.Series(x, name="thalach variable")
ax = sns.kdeplot(x)
plt.show()
```



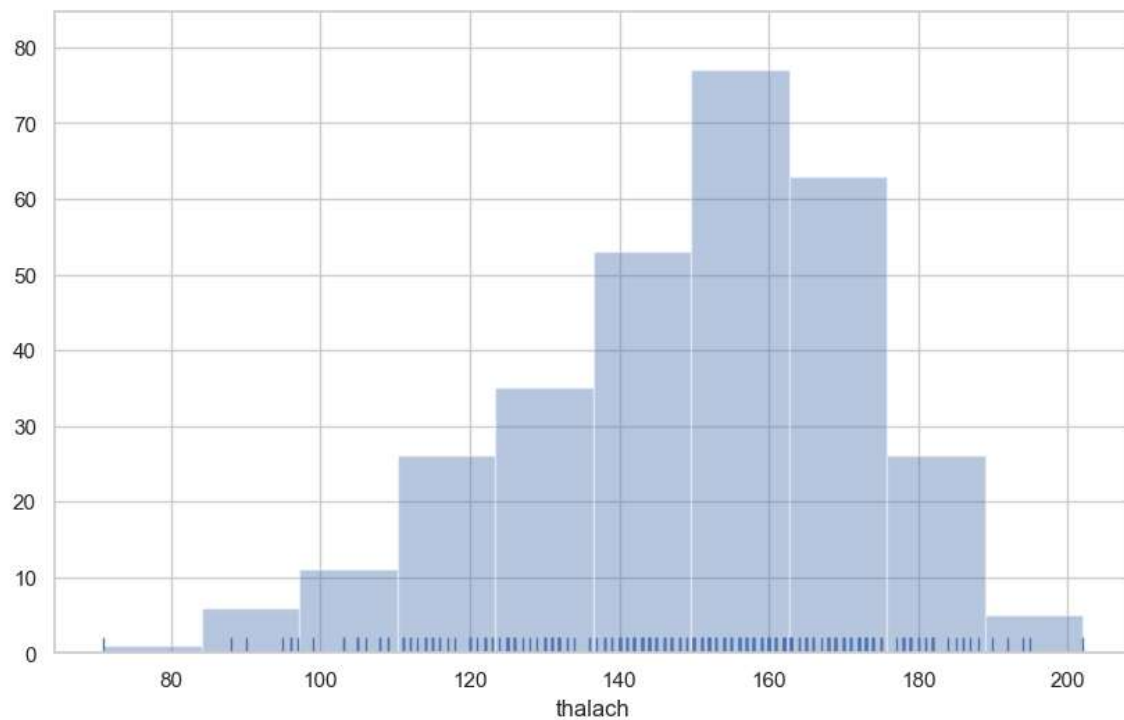
In [36]:

```
f, vis = plt.subplots(figsize=(10,6))
x = heart['thalach']
x = pd.Series(x, name="thalach variable")
ax = sns.kdeplot(x, shade=True, color='r')
plt.show()
```



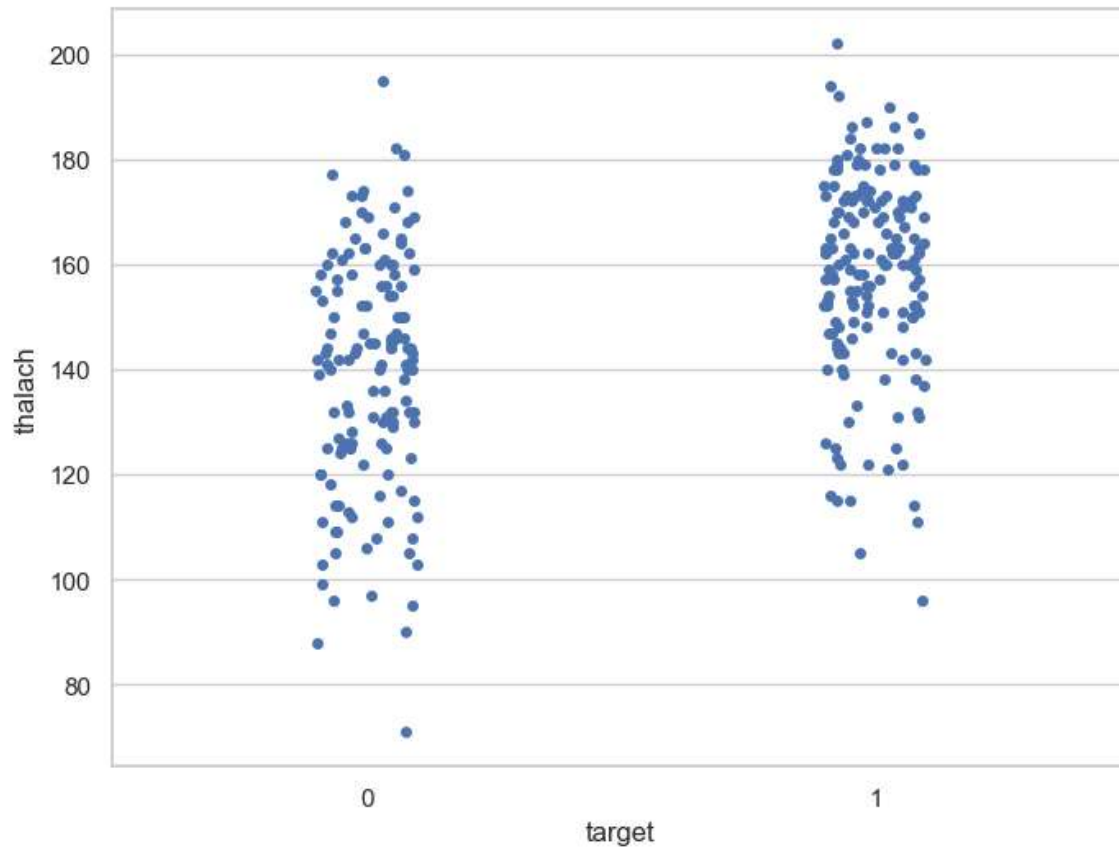
In [37]:

```
f, vis = plt.subplots(figsize=(10,6))
x = heart['thalach']
ax = sns.distplot(x, kde=False, rug=True, bins=10)
plt.show()
```



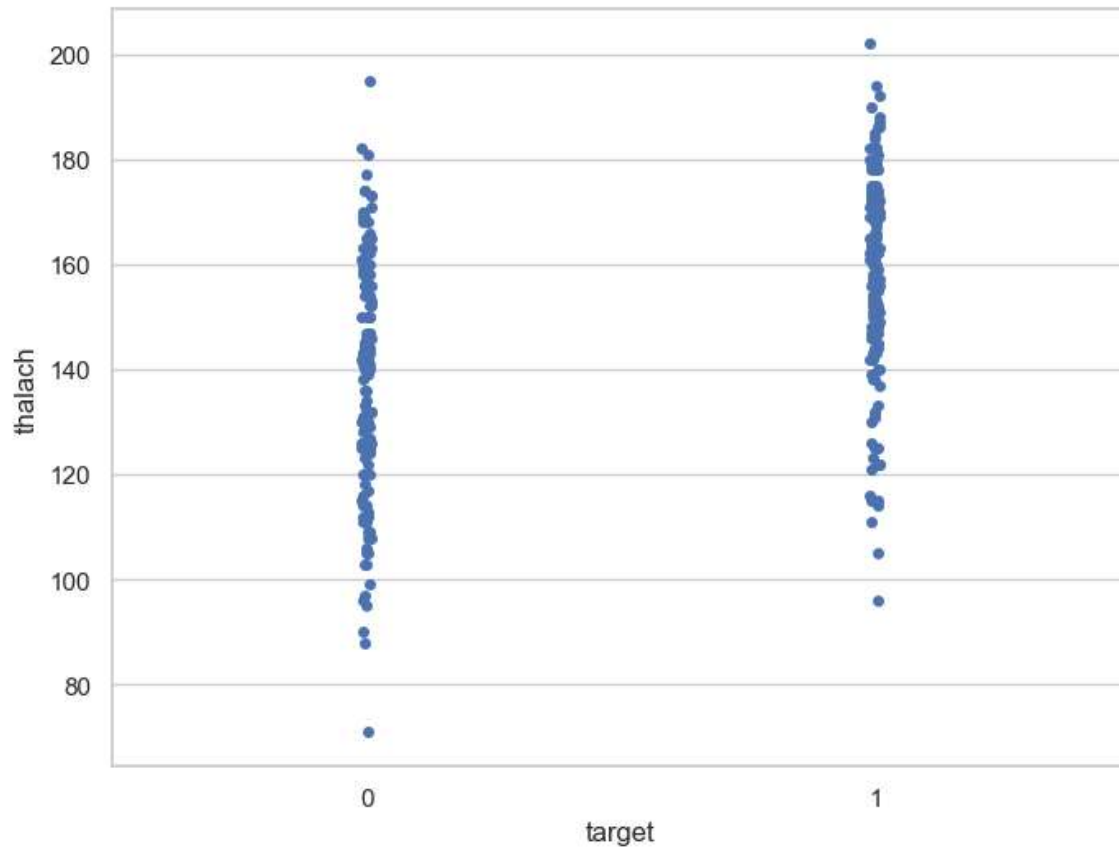
In [38]:

```
f, vis = plt.subplots(figsize=(8, 6))  
sns.stripplot(x="target", y="thalach", data=heart)  
plt.show()
```



In [39]:

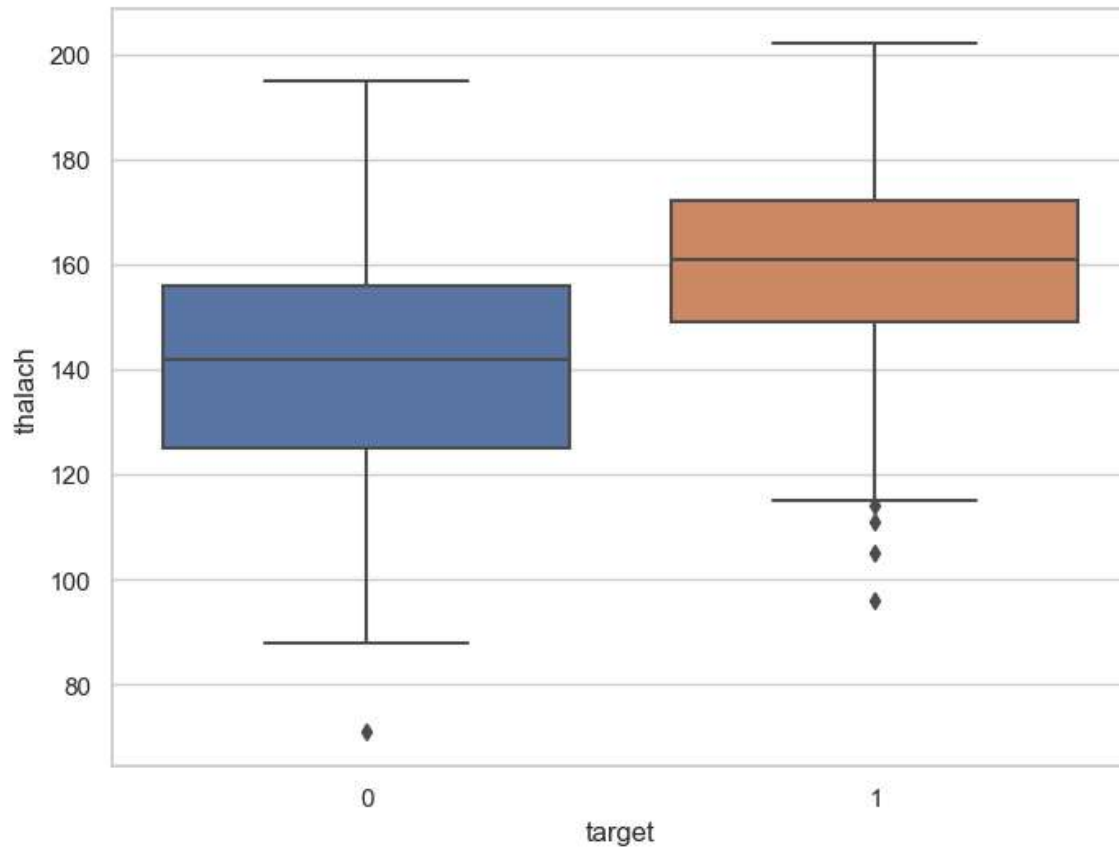
```
f, vis = plt.subplots(figsize=(8, 6))  
sns.stripplot(x="target", y="thalach", data=heart, jitter = 0.01)  
plt.show()
```





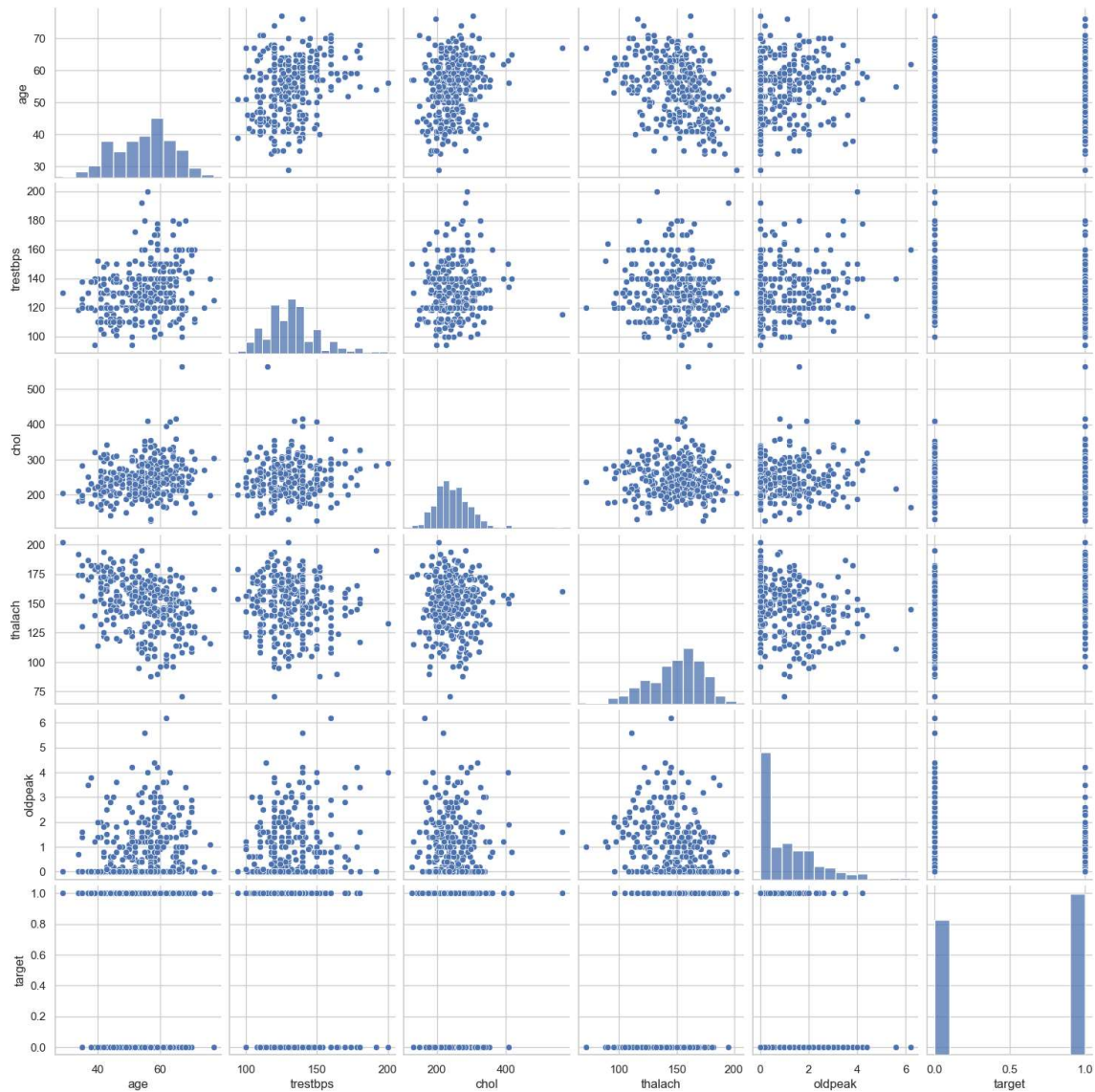
In [40]:

```
f, vis = plt.subplots(figsize=(8, 6))  
sns.boxplot(x="target", y="thalach", data=heart)  
plt.show()
```



In [41]:

```
num_var = ['age', 'trestbps', 'chol', 'thalach', 'oldpeak', 'target']  
sns.pairplot(heart[num_var], kind='scatter', diag_kind='hist')  
plt.show()
```



In [43]:

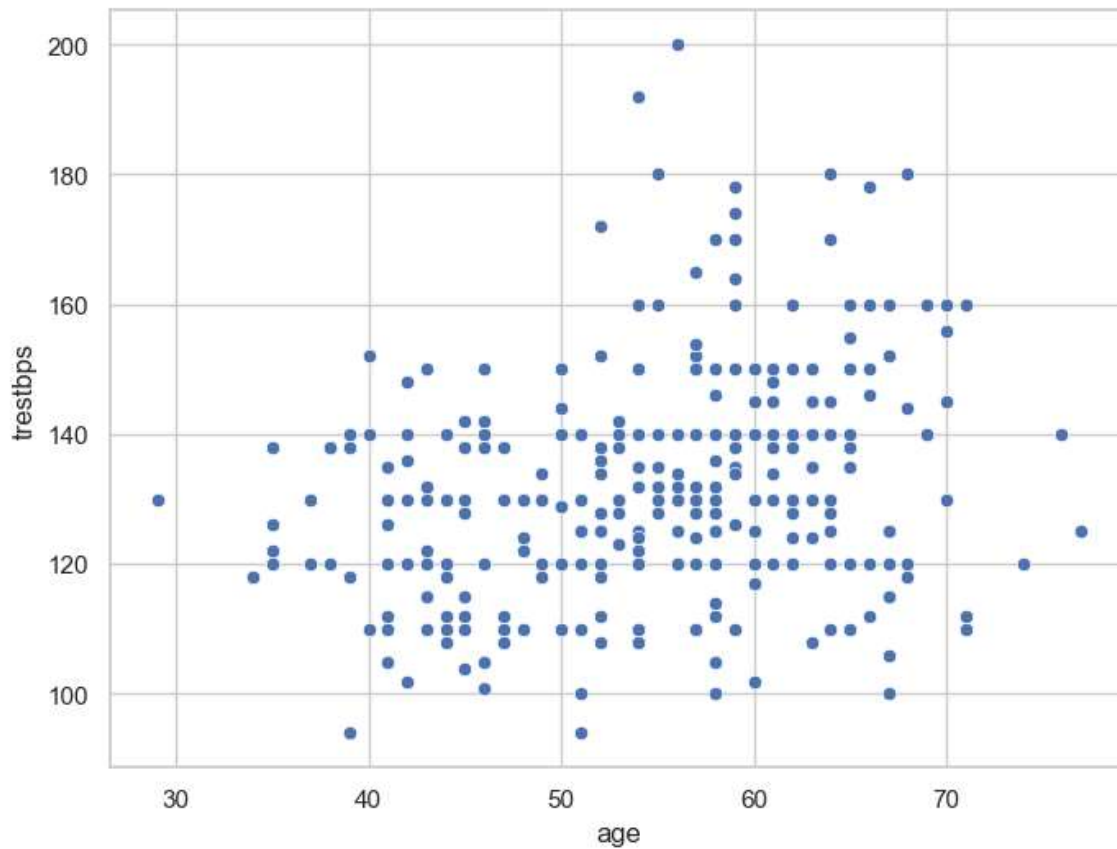
```
num_var
```

Out[43]:

```
['age', 'trestbps', 'chol', 'thalach', 'oldpeak', 'target']
```

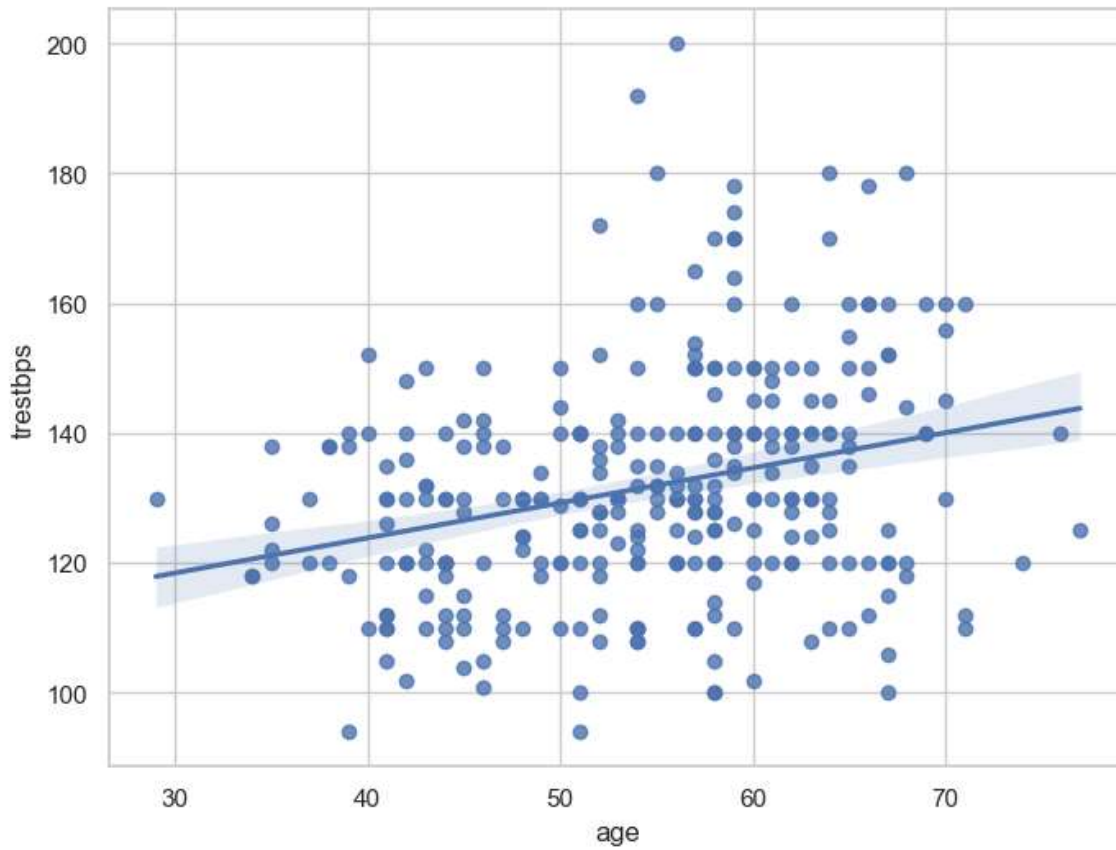
In [44]:

```
f, vis = plt.subplots(figsize=(8, 6))  
ax = sns.scatterplot(x="age", y="trestbps", data=heart)  
plt.show()
```



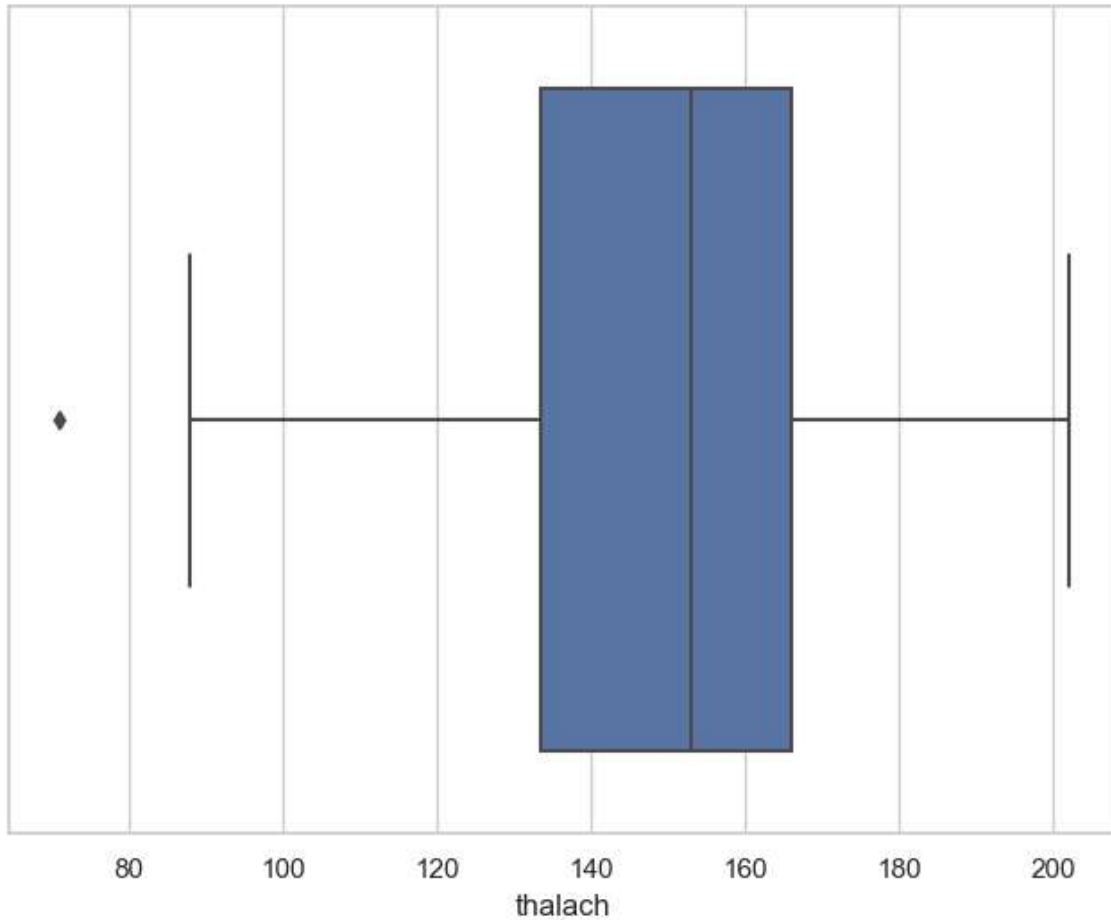
In [45]:

```
f, vis = plt.subplots(figsize=(8, 6))  
ax = sns.regplot(x="age", y="trestbps", data=heart)  
plt.show()
```



In [46]:

```
f, ax = plt.subplots(figsize=(8, 6))  
sns.boxplot(x=heart["thalach"])  
plt.show()
```



In [47]:

```
heart['oldpeak'].describe()
```

Out[47]:

```
count    303.000000  
mean      1.039604  
std       1.161075  
min       0.000000  
25%       0.000000  
50%       0.800000  
75%       1.600000  
max       6.200000  
Name: oldpeak, dtype: float64
```

In [48]:

```
f, vis= plt.subplots(figsize=(8, 6))  
sns.boxplot(x=heart["oldpeak"])  
plt.show()
```

