```
import pandas as pd

df = pd.read_csv('heart.csv')
```

Note: Here in Dataset "trtbps" assume it as RestBP

```
df
# 303 rows × 14 columns
```

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

     caa  thall  output
0      0      1       1
1      0      2       1
2      0      2       1
3      0      2       1
4      0      2       1
..   ...    ...     ...
298    0      3       0
299    0      3       0
300    2      3       0
301    1      3       0
302    1      2       0

[303 rows x 14 columns]
```

```
df = df.drop_duplicates()
```

```
df
```

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

     caa  thall  output
0      0      1       1
1      0      2       1
2      0      2       1
3      0      2       1
4      0      2       1
..   ...    ...     ...
298    0      3       0
299    0      3       0
300    2      3       0
301    1      3       0
302    1      2       0

[302 rows x 14 columns]
```

Duplicates removed

```
df.isna().sum()
# No null values, it's clean
```

```
age         0
sex         0
cp          0
```

```
trtbps      0
chol        0
fbs         0
restecg     0
thalachh    0
exng        0
oldpeak     0
slp         0
caa         0
thall       0
output      0
dtype: int64
```

Plots

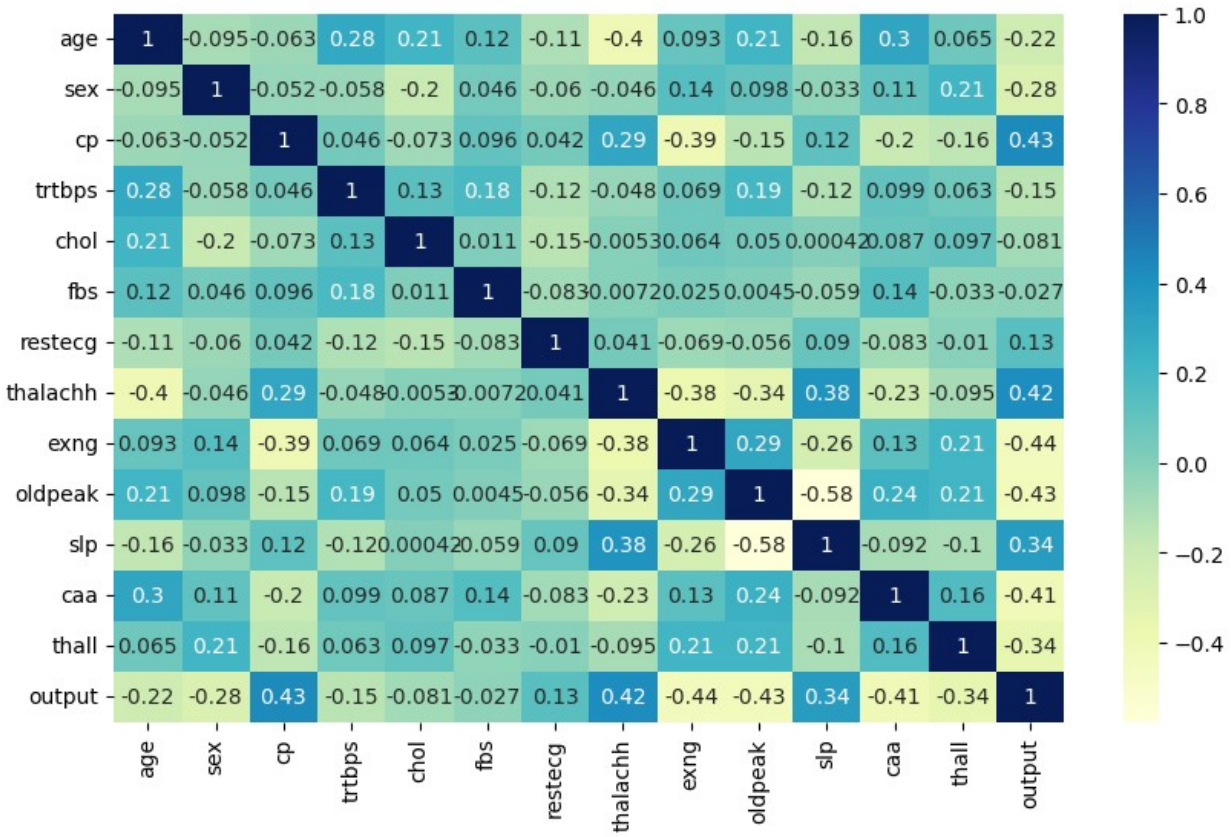
```
import seaborn as sns
import matplotlib.pyplot as plt

df.columns
```

```
Index(['age', 'sex', 'cp', 'trtbps', 'chol', 'fbs', 'restecg',
'thalachh',
       'exng', 'oldpeak', 'slp', 'caa', 'thall', 'output'],
      dtype='object')
```

```
plt.figure(figsize=(10,6))
sns.heatmap(df.corr(),cmap = 'YlGnBu', annot = True)
```

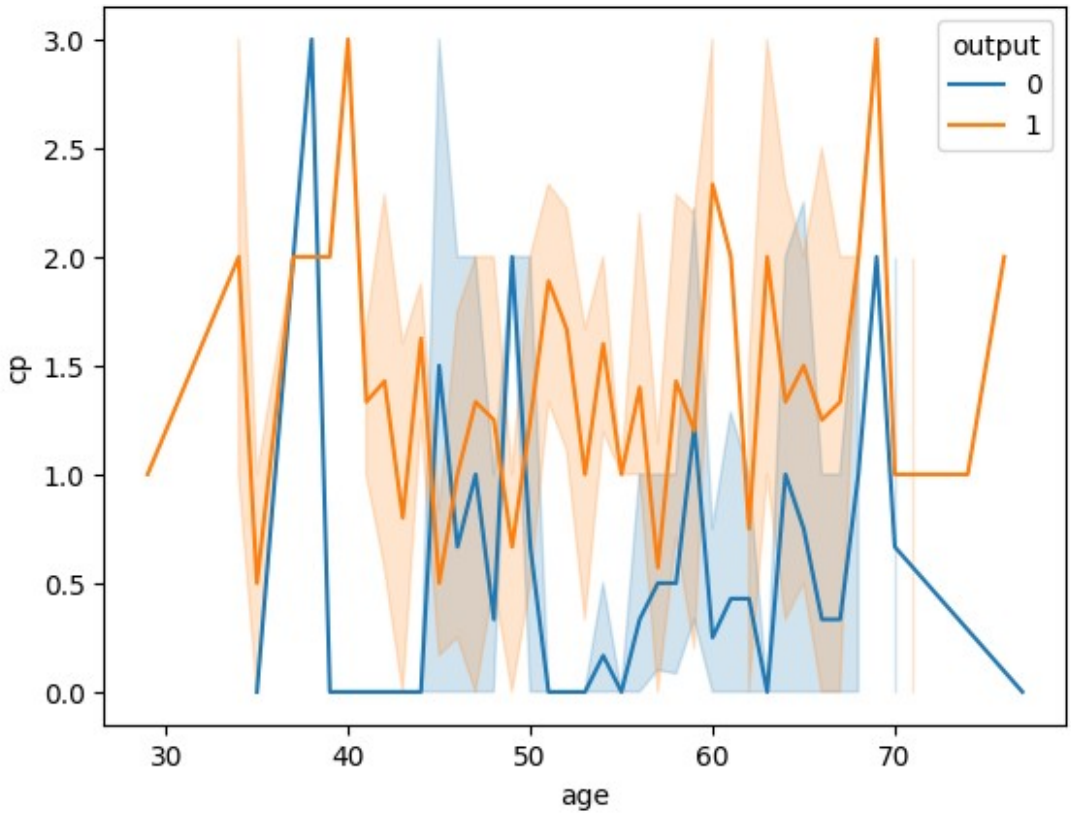
```
<Axes: >
```



Line chart

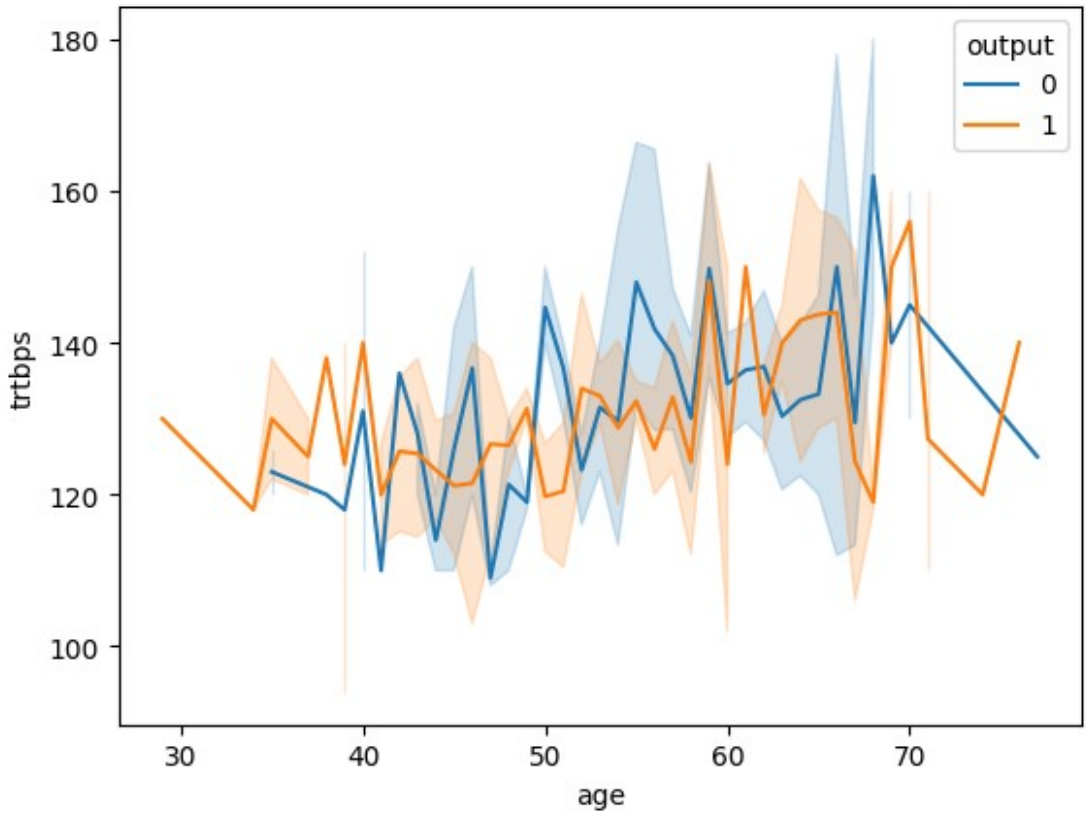
```
sns.lineplot(data=df,x=df.age,y=df.cp,hue='output')
```

```
<Axes: xlabel='age', ylabel='cp'>
```



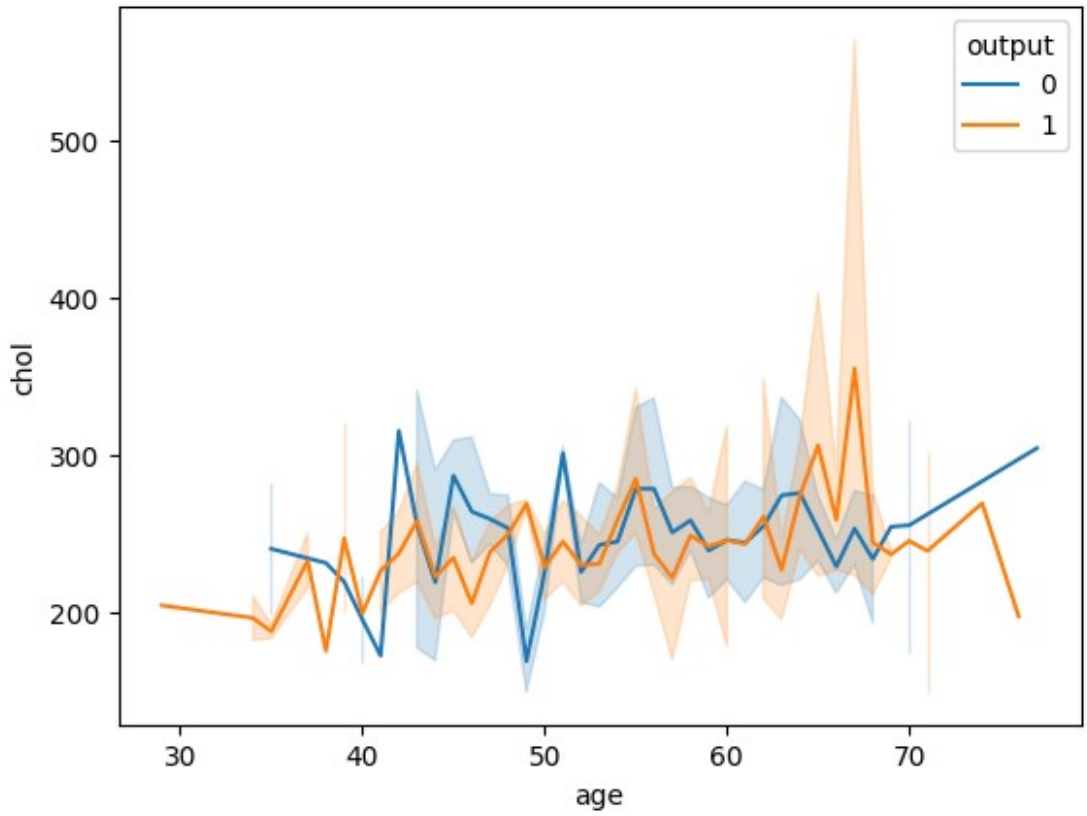

```
sns.lineplot(data=df,x=df.age,y=df.trtbps,hue='output')
```

```
<Axes: xlabel='age', ylabel='trtbps'>
```



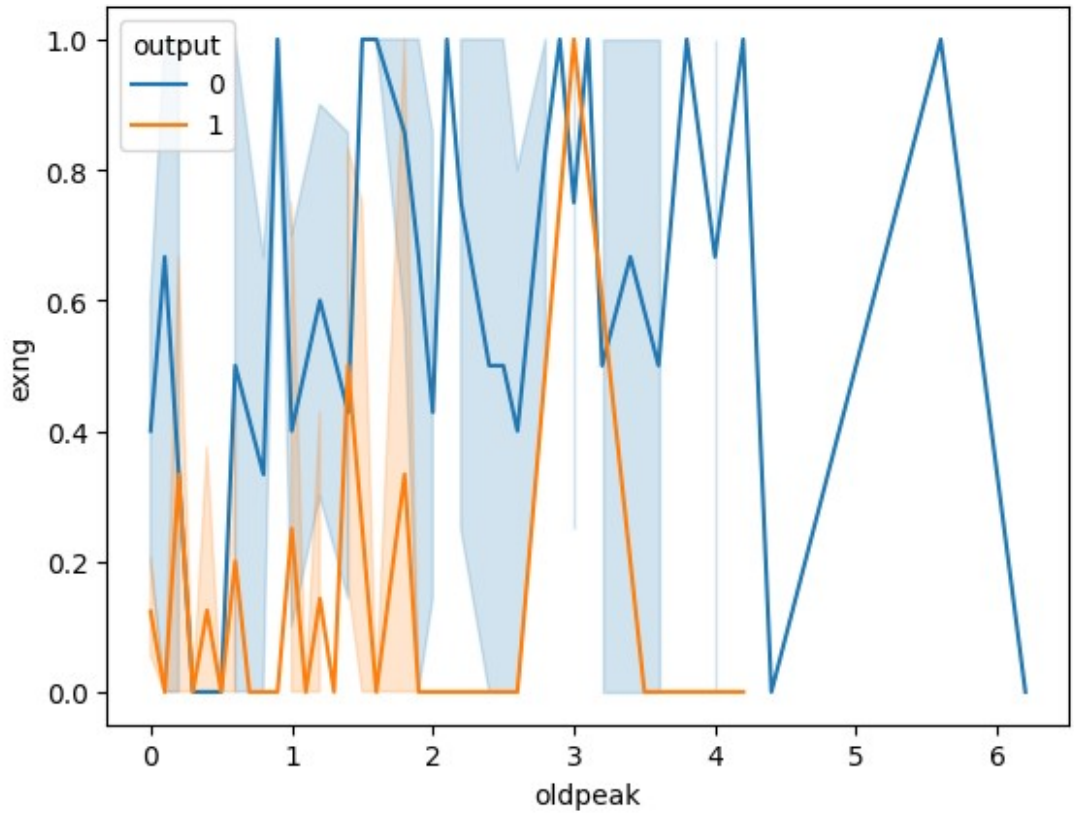

```
sns.lineplot(data=df,x=df.age,y=df.chol,hue='output')
```

```
<Axes: xlabel='age', ylabel='chol'>
```



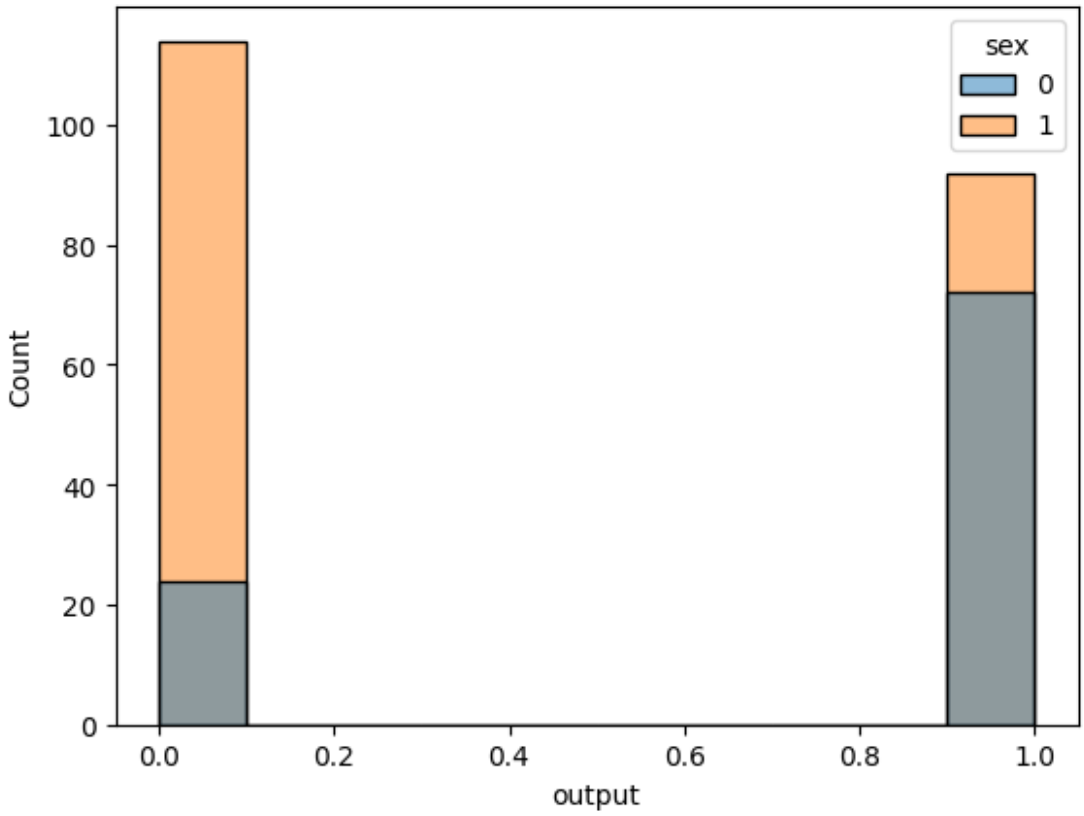

```
sns.lineplot(df,x=df.oldpeak,y=df.exng,hue='output')
```

```
<Axes: xlabel='oldpeak', ylabel='exng'>
```



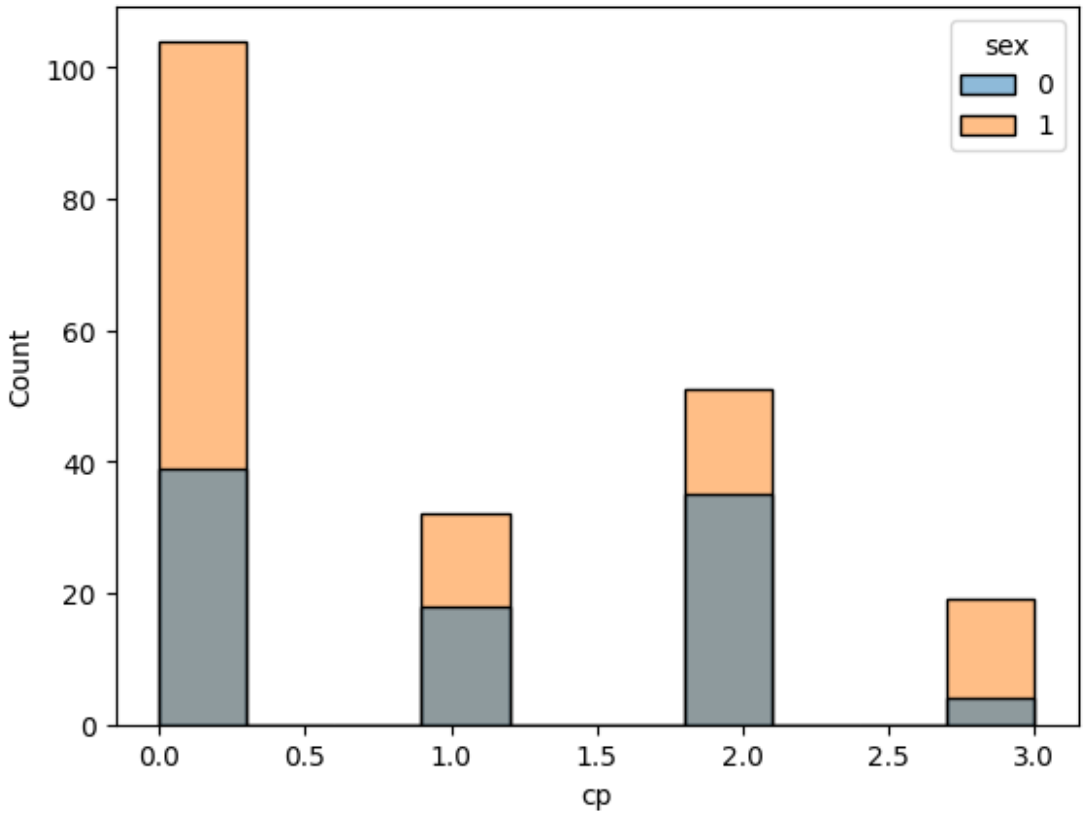

```
# Shows the Distribution of Heat Diseases with respect to male and
female
sns.histplot(data=df,
             x=df.output,
             hue=df.sex)
```

```
<Axes: xlabel='output', ylabel='Count'>
```



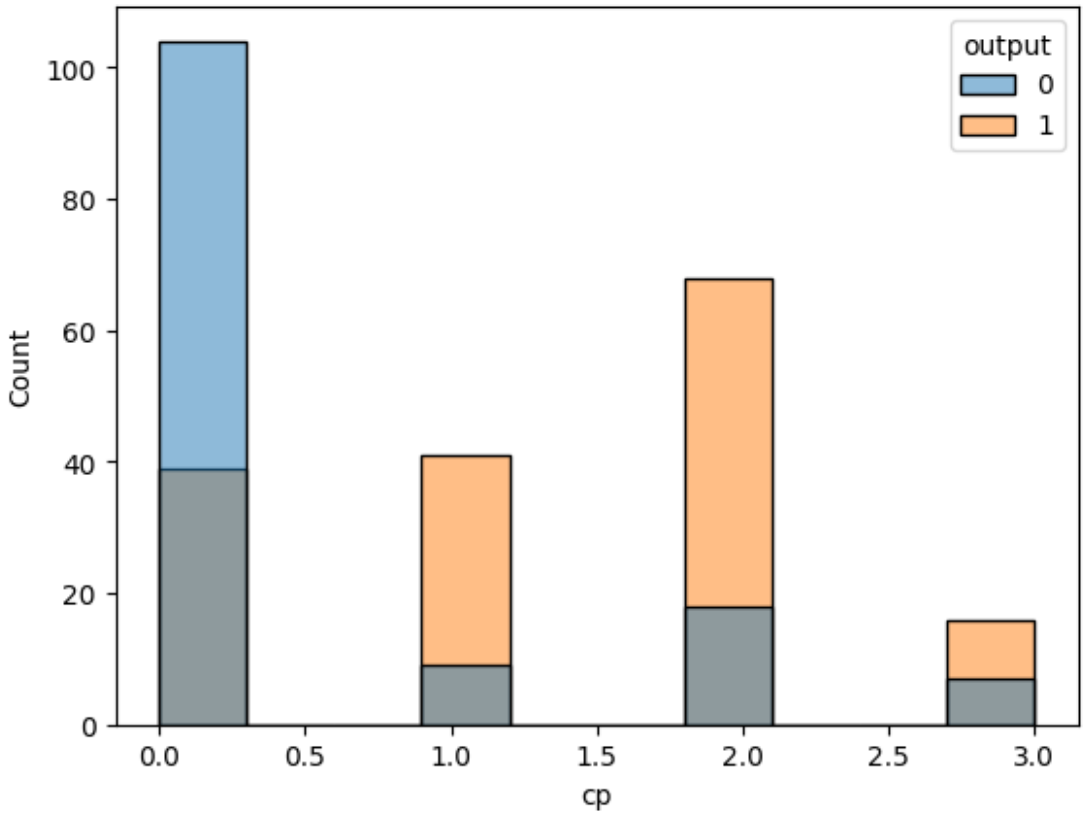

```
# Shows the Distribution of cp types with respect to male and female
sns.histplot(data=df,
             x=df.cp,
             hue=df.sex)
```

```
<Axes: xlabel='cp', ylabel='Count'>
```



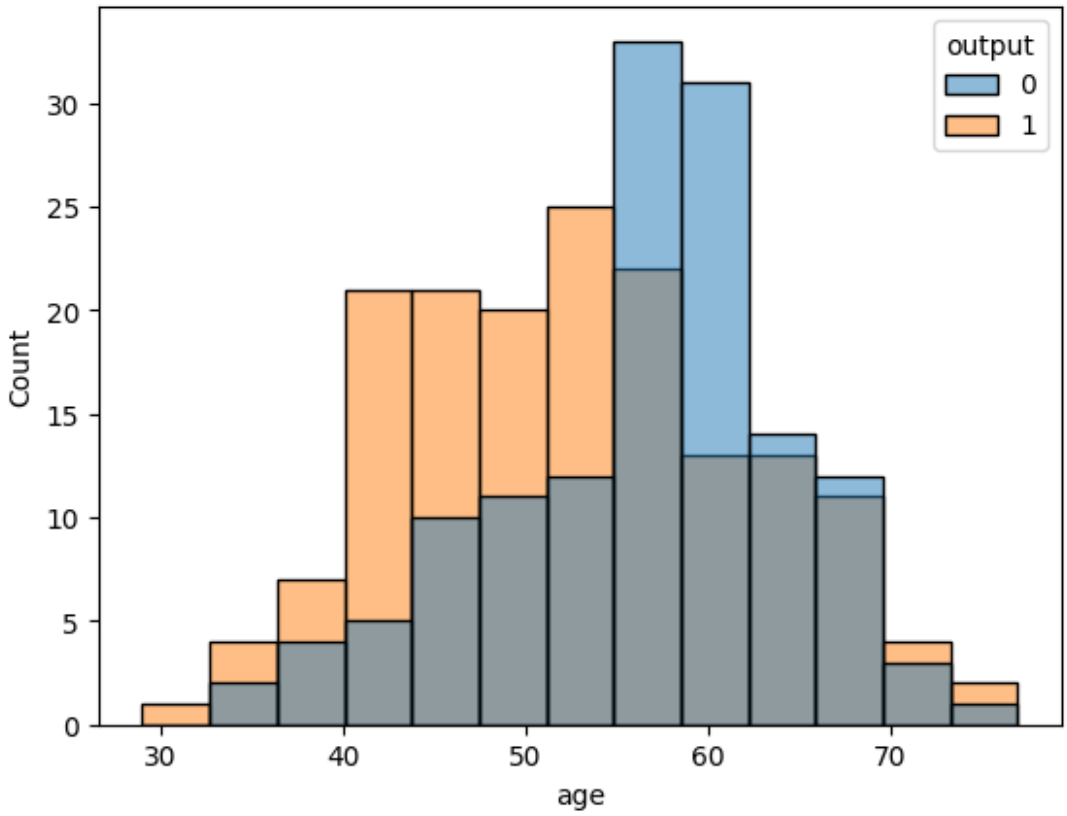

```
sns.histplot(data=df,x=df.cp, hue='output')
```

```
<Axes: xlabel='cp', ylabel='Count'>
```



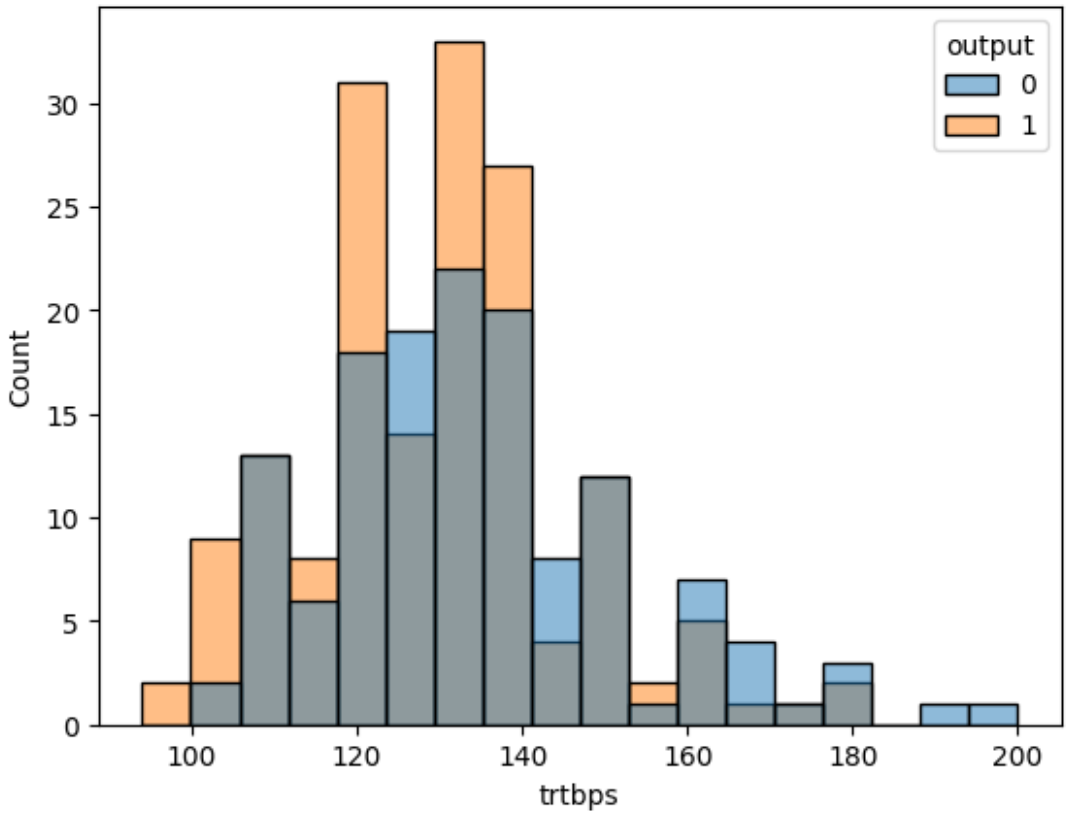

```
# Shows the Distribution of age w.r.t output
sns.histplot(data=df,x=df['age'], hue='output')
```

```
<Axes: xlabel='age', ylabel='Count'>
```




```
sns.histplot(data=df,x=df.trtbps, hue='output')
```

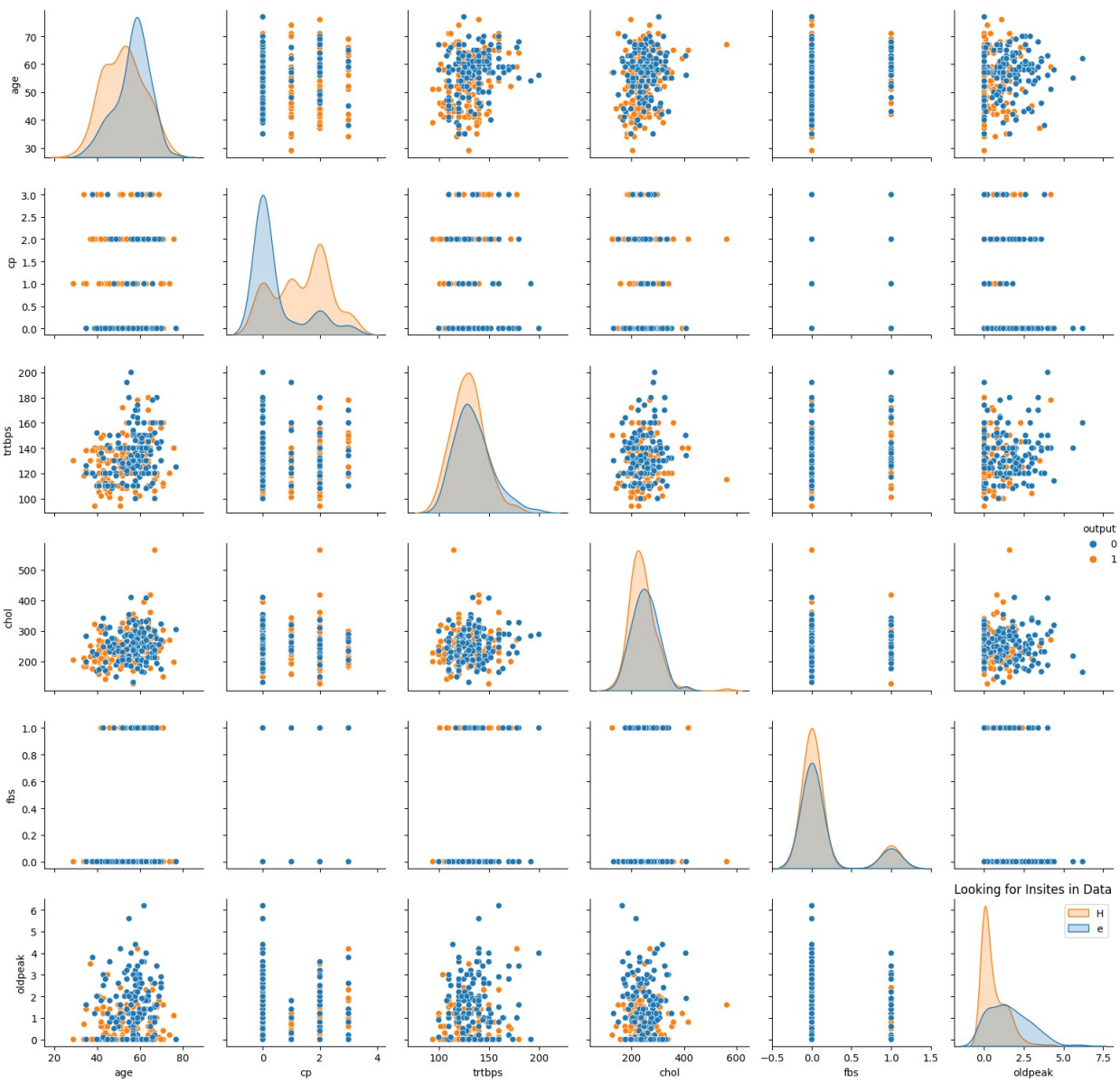
```
<Axes: xlabel='trtbps', ylabel='Count'>
```



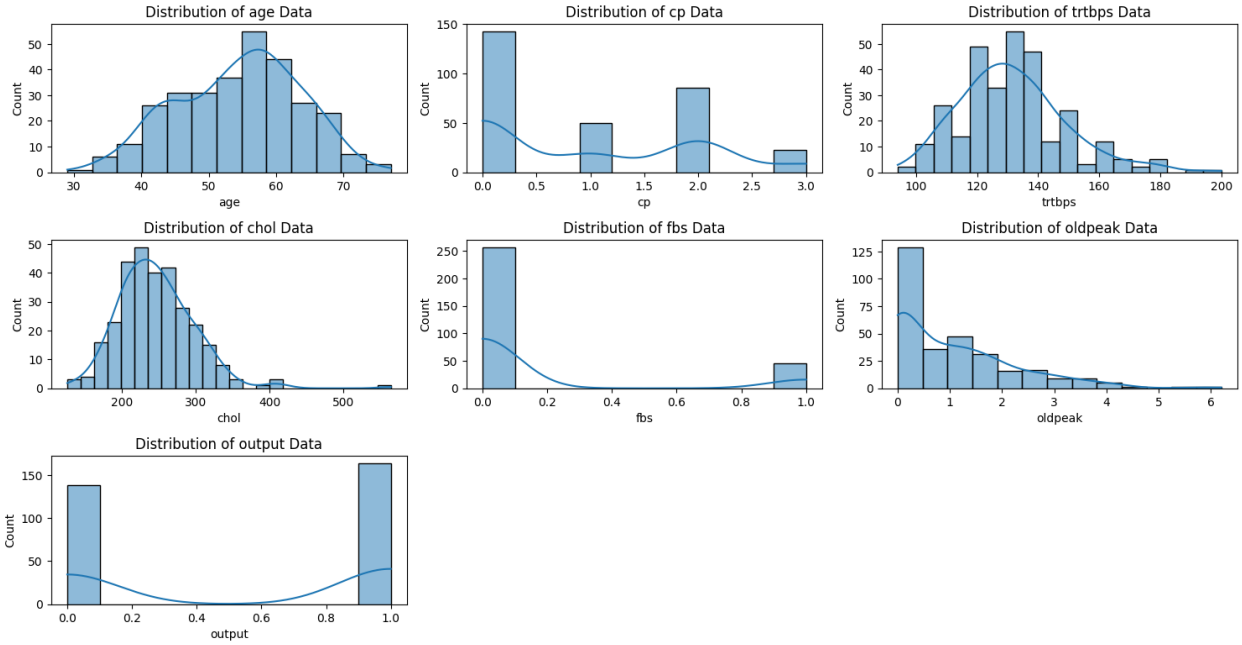

```
temp_df = df[['age','cp', 'trtbps','chol','fbs','oldpeak','output']]
plt.figure(figsize=(15,10))
sns.pairplot(temp_df,hue="output")
plt.title("Looking for Insites in Data")
plt.legend("HeartDisease")
plt.tight_layout()
plt.plot()
```

```
[]
```

```
<Figure size 1500x1000 with 0 Axes>
```




```
plt.figure(figsize=(15,10))
for i,col in enumerate(temp_df.columns,1):
    plt.subplot(4,3,i)
    plt.title(f"Distribution of {col} Data")
    sns.histplot(df[col],kde=True)
    plt.tight_layout()
    plt.plot()
```


Distribution of age Data
Count
age
Distribution of cp Data
Count
cp
Distribution of trtbps Data
Count
trtbps
Distribution of chol Data
Count
chol
Distribution of fbs Data
Count
fbs
Distribution of oldpeak Data
Count
oldpeak
Distribution of output Data
Count
output