讀檔 cancer.csv，import

```python
import pandas as pd
pd.plotting.register_matplotlib_converters()
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
print("Setup Complete")
```

Setup Complete

```python
# Set up code checking
from learntools.core import binder
binder.bind(globals())
from learntools.data_viz_to_coder.ex5 import *
print("Setup Complete")
```

Setup Complete

```python
# Path of the files to read
cancer_filepath = "../input/cancer.csv"

# Fill in the line below to read the file into a variable cancer_data
cancer_data = pd.read_csv(cancer_filepath, index_col="Id")

# Run the line below with no changes to check that you've loaded the data correctly
step_1.check()
```

Step2:Print 前五筆資料

```python
# Print the first five rows of the data
cancer_data.head() # Your code here
```

Id	Diagnosis	Radius (mean)	Texture (mean)	Perimeter (mean)	Area (mean)	Smoothness (mean)	Compactness (mean)	Concavity (mean)	Concave points (mean)	Symmetry (mean)	...	Radius (worst)	Texture (worst)	Perimeter (worst)	Are (wors
8510426	B	13.540	14.36	87.46	566.3	0.09779	0.08129	0.06664	0.047810	0.1885	...	15.110	19.26	99.70	711
8510653	B	13.080	15.71	85.63	520.0	0.10750	0.12700	0.04568	0.031100	0.1967	...	14.500	20.49	96.09	630
8510824	B	9.504	12.44	60.34	273.9	0.10240	0.06492	0.02956	0.020760	0.1815	...	10.230	15.66	65.13	314
854941	B	13.030	18.42	82.61	523.8	0.08983	0.03766	0.02562	0.029230	0.1467	...	13.300	22.81	84.46	545
85713702	B	8.196	16.84	51.71	201.9	0.08600	0.05943	0.01588	0.005917	0.1769	...	8.964	21.96	57.26	242

5 rows × 31 columns

Step3:使用下方代碼單元建立兩個直方圖，分別顯示良性和惡性腫瘤的「面積（平均值）」值的分佈。

```
# Histograms for benign and maligant tumors
sns.histplot(data=cancer_data, x='Area (mean)', hue='Diagnosis')

# Check your answer
step_3.a.check()
```
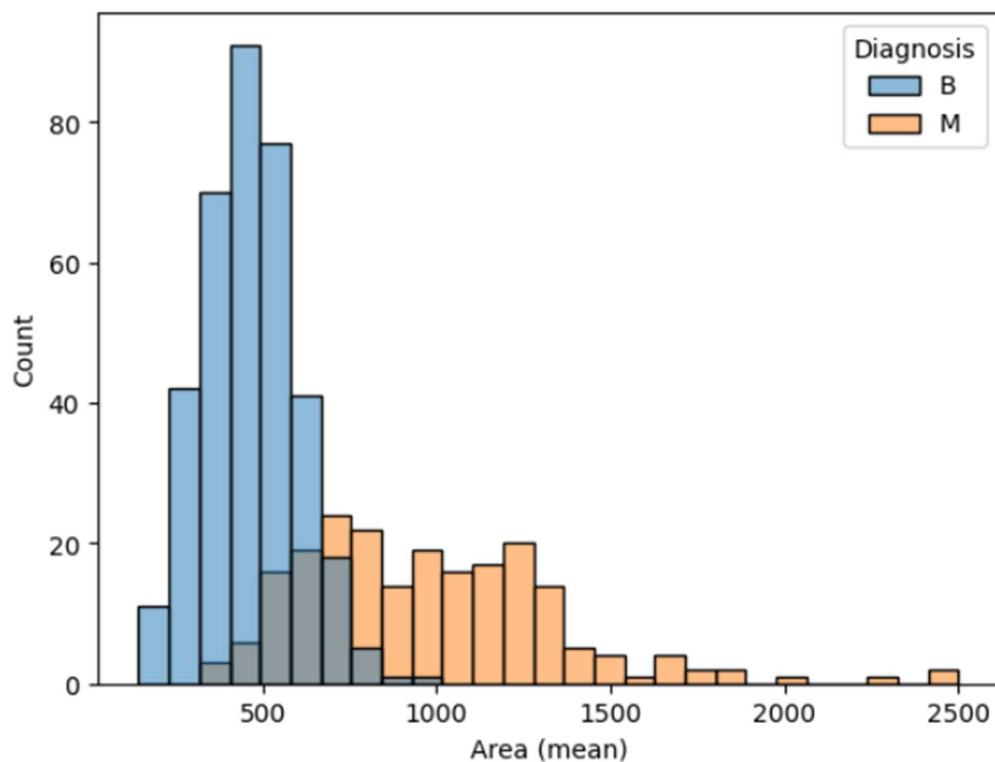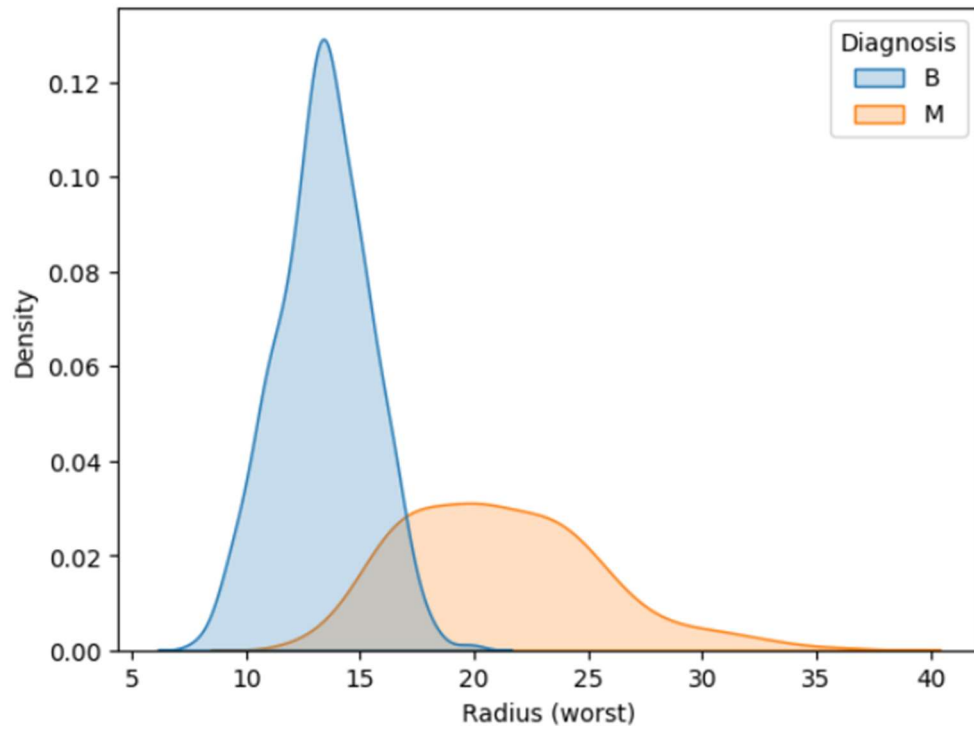

惡性腫瘤的「面積（平均值）」平均數值較高。惡性腫瘤的潛在值範圍較大。

Step4:使用下方代碼單元建立兩個 KDE 圖，分別顯示良性和惡性腫瘤的「半徑（最差）」值的分佈。（為了方便比較，請在下方程式碼單元中建立一個包含兩個 KDE 圖的圖形。）

```
# KDE plots for benign and malignant tumors
sns.kdeplot(data=cancer_data, x='Radius (worst)', hue='Diagnosis', shade=True)

# Check your answer
step_4.a.check()
```

該演算法更有可能將腫瘤歸類為惡性腫瘤。這是因為惡性腫瘤的曲線遠高於良性腫瘤的曲線，其值在 25 左右——而準確率高的演算法很可能會根據數據中的這種模式做出決策。