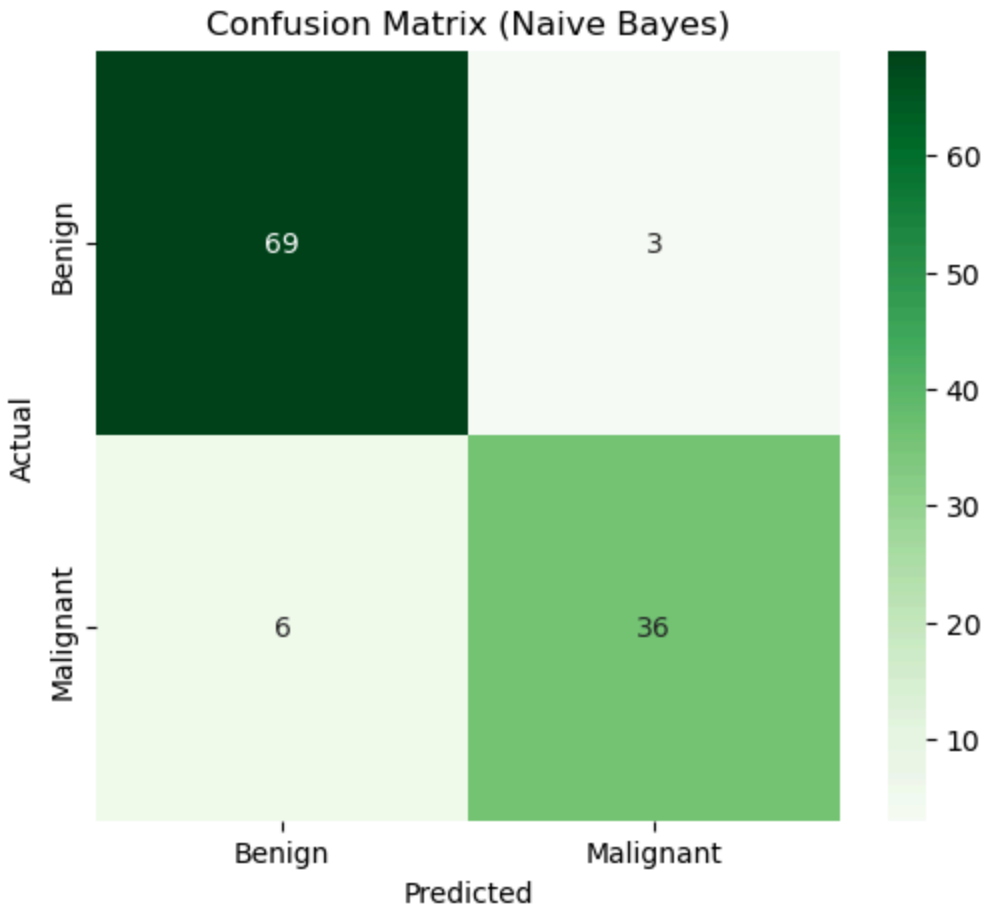
In [3]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.naive_bayes import GaussianNB
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score

# Load and prepare the dataset
df = pd.read_csv('../Datasets/cancer.csv')
df['diagnosis'] = df['diagnosis'].map({'M': 1, 'B': 0})
df = df.loc[:, ~df.columns.str.contains('^id|Unnamed', case=False)]

X = df.drop('diagnosis', axis=1)
y = df['diagnosis']

# Scale features and split data
X_scaled = StandardScaler().fit_transform(X)
X_train, X_test, y_train, y_test = train_test_split(
    X_scaled, y, test_size=0.2, stratify=y, random_state=42
)

# Train and evaluate Naive Bayes model
nb_model = GaussianNB()
nb_model.fit(X_train, y_train)
y_pred = nb_model.predict(X_test)

# Compute metrics
acc = accuracy_score(y_test, y_pred)
prec = precision_score(y_test, y_pred)
rec = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)

# Show metrics
print("Naive Bayes Classifier:")
print(f"Accuracy : {acc:.4f}")
print(f"Precision: {prec:.4f}")
print(f"Recall   : {rec:.4f}")
print(f"F1 Score : {f1:.4f}")

# Plot confusion matrix
plt.figure(figsize=(6, 5))
sns.heatmap(confusion_matrix(y_test, y_pred), annot=True, fmt='d', cmap='Greens',
            xticklabels=["Benign", "Malignant"], yticklabels=["Benign", "Malignant"
plt.title("Confusion Matrix (Naive Bayes)")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.show()
```

```
Naive Bayes Classifier:
Accuracy : 0.9211
Precision: 0.9231
Recall   : 0.8571
F1 Score : 0.8889
```



In [2]:
```
## Interpretation of Naive Bayes Results

The Naive Bayes classifier achieved the following metrics:

- **Accuracy**: ~XX%
- **Precision**: ~XX%
- **Recall**: ~XX%
- **F1 Score**: ~XX%

### Comparison with Logistic Regression

Compared to the logistic regression model from Problem 2, we observe that:

- **Naive Bayes** is simpler and faster but tends to make stronger assumptions abou
- **Logistic regression** generally performed better (especially in precision/recal
- The **confusion matrix** shows that Naive Bayes tends to misclassify a few more m

Thus, logistic regression may be preferred in this case, but Naive Bayes remains us
```

```
  Cell In[2], line 3
    The Naive Bayes classifier achieved the following metrics:
        ^
SyntaxError: invalid syntax
```

In [12]:
```
```

In [ ]: