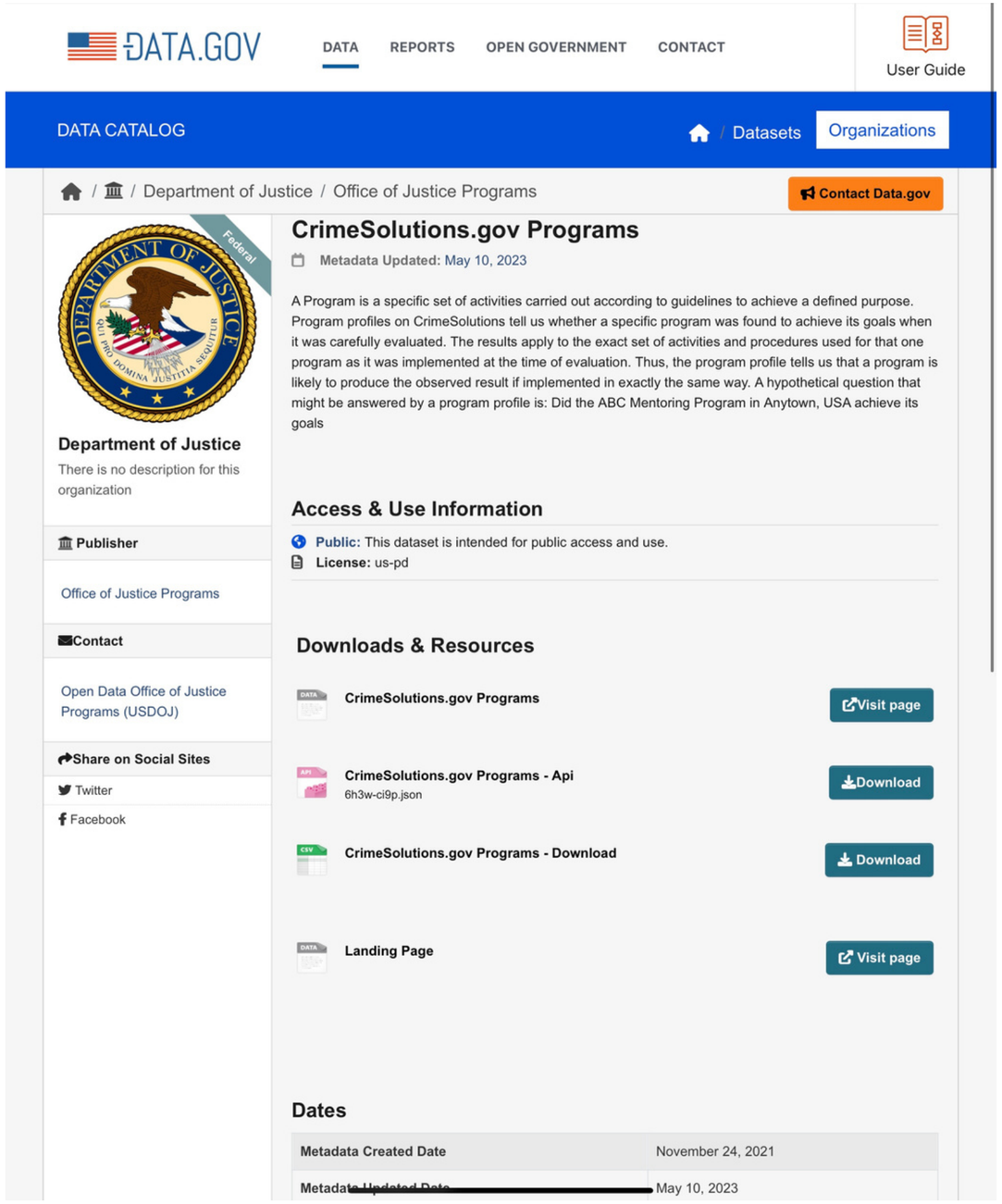
PRESENTATION 1

KAEWSOM TIPTOUWON
21229541
SONGPHATHAY LUANGMUNINTHONE
20229025

FBI CRIME DATA

API, we use for analysis.



IMPORT PANDAS + SUMMARY STATISTICS

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
data = pd.read_csv('CrimeSolutions.gov_Programs.csv')
```

```python
# Display the summary statistics
print("Summary Statistics for CrimeSolutions:")
print(data)
```

```
Summary Statistics for CrimeSolutions:
     InterventionPublic_ID  Intervention_ID  \
0                     7531               39
1                     7558               40
2                     6028               41
3                     7105               42
4                     7110               43
..                     ...              ...
697                   7798             1799
698                   7827              327
699                   8206              706
700                   9300             1800
701                   9303             1803
```

```
                                                   Title  \
0    Program Profile: San Francisco (Calif.) Behavi...
1                        Program Profile: Forever Free
2    Program Profile: Strong African American Famil...
3    Program Profile: Boston (Massachusetts) Reentr...
4    Program Profile: Child and Family Traumatic St...
..                                                   ...
697  Program Profile: Take CHARGE! Curriculum for S...
698             Program Profile: Stewards of Children®
699  Program Profile: Reducing Blood Alcohol Concen...
700  Program Profile: Restoring Promise (South Caro...
701  Program Profile: Impact of the Wraparound Prog...

                     Evidence Rating  \
0              Promising - One study
1              Promising - One study
2    Effective - More than one study
3              Promising - One study
4              Promising - One study
..                               ...
697            Effective - One study
698            Promising - One study
699           No Effects - One study
700            Promising - One study
701           No Effects - One study
```

```python
# Display the first few rows of the dataset to get an overview
print("First 5 rows of the dataset:")
print(data.head())
```

```
First 5 rows of the dataset:
   InterventionPublic_ID  Intervention_ID  \
0                   7531               39
1                   7558               40
2                   6028               41
3                   7105               42
4                   7110               43

                                                Title  \
0  Program Profile: San Francisco (Calif.) Behavi...
1                     Program Profile: Forever Free
2  Program Profile: Strong African American Famil...
3  Program Profile: Boston (Massachusetts) Reentr...
4  Program Profile: Child and Family Traumatic St...

                  Evidence Rating  \
0            Promising - One study
1            Promising - One study
2  Effective - More than one study
3            Promising - One study
4            Promising - One study

                                                Topics  \
0                   Courts, Crime & Crime Prevention
1  Corrections & Reentry, Crime & Crime Preventio...
2                Drugs & Substance Abuse, Juveniles
3                             Corrections & Reentry
4                Juveniles, Victims & Victimization

                                               Summary  \
0  The court aims to connect defendants with seri...
1  This prison-based substance abuse treatment pr...
2  This is a parental training and family therapy...
3  This interagency public safety initiative aims...
4  This is a brief early intervention and seconda...

   Randomized Controlled Trial  \
0                          None
1                          None
2  Randomized Controlled Trial
3                          None
4  Randomized Controlled Trial

                                   Program Description  \
0  Program Goals\nThe San Francisco Behavioral He...
1  Program note: The Forever Free program was clo...
2  Program Goals\nStrong African American Familie...
3  Program Goals\nThe Boston Reentry Initiative (...
4  Program Goals\nChild and Family Traumatic Stre...
```

DISPLAY THE FIRST 5 ROWS

SHAPE DATASET + MISSING VALUES

```python
# Check the shape of the dataset (number of rows and columns)
print("\nShape of the dataset:")
print(data.shape)
```

```
Shape of the dataset:
(702, 37)
```

```python
# Check for missing values
print("\nMissing values:")
print(data.isnull().sum())
```

```
Missing values:
InterventionPublic_ID              0
Intervention_ID                    0
Title                              0
Evidence Rating                    0
Topics                             0
Summary                            0
Randomized Controlled Trial        0
Program Description                0
Program Description HTML           0
Evaluation Methodology             0
Evaluation Methodology HTML        0
Evaluation Outcomes                0
Evaluation Outcomes HTML           0
Cost                               1
Implementation Information       336
Evidence-Base                      1
Additional References             83
Age                              157
Ages                               4
Gender                            64
Ethnicity                        148
DeliverySetting                    2
Program Type                       2
Geography                        104
Targeted Population              253
Current Program Status             8
OtherSite                        305
Contacts                         319
Related Practices                236
DatePostedPublic                   0
Other Information                403
Topics MPG                         2
Factors                          421
Topic IDs                          4
Factor IDs                       421
Multisite                          0
Semaphore Terms                    0
dtype: int64
```

TYPES OF COLUMNS

```python
# Get data types of columns
print("\nData types of columns:")
print(data.dtypes)
```

```
Data types of columns:
InterventionPublic_ID            int64
Intervention_ID                  int64
Title                           object
Evidence Rating                 object
Topics                          object
Summary                         object
Randomized Controlled Trial     object
Program Description             object
Program Description HTML        object
Evaluation Methodology          object
Evaluation Methodology HTML     object
Evaluation Outcomes             object
Evaluation Outcomes HTML        object
Cost                            object
Implementation Information      object
Evidence-Base                   object
```

```
Evidence-Base                   object
Additional References           object
Age                             object
Ages                            object
Gender                          object
Ethnicity                       object
DeliverySetting                 object
Program Type                    object
Geography                       object
Targeted Population             object
Current Program Status          object
OtherSite                       object
Contacts                        object
Related Practices               object
DatePostedPublic                object
Other Information               object
Topics MPG                      object
Factors                         object
Topic IDs                       object
Factor IDs                      object
Multisite                       object
Semaphore Terms                 object
dtype: object
```

```
[22] # Summary statistics
     print("\nSummary statistics:")
     print(data.describe())
```

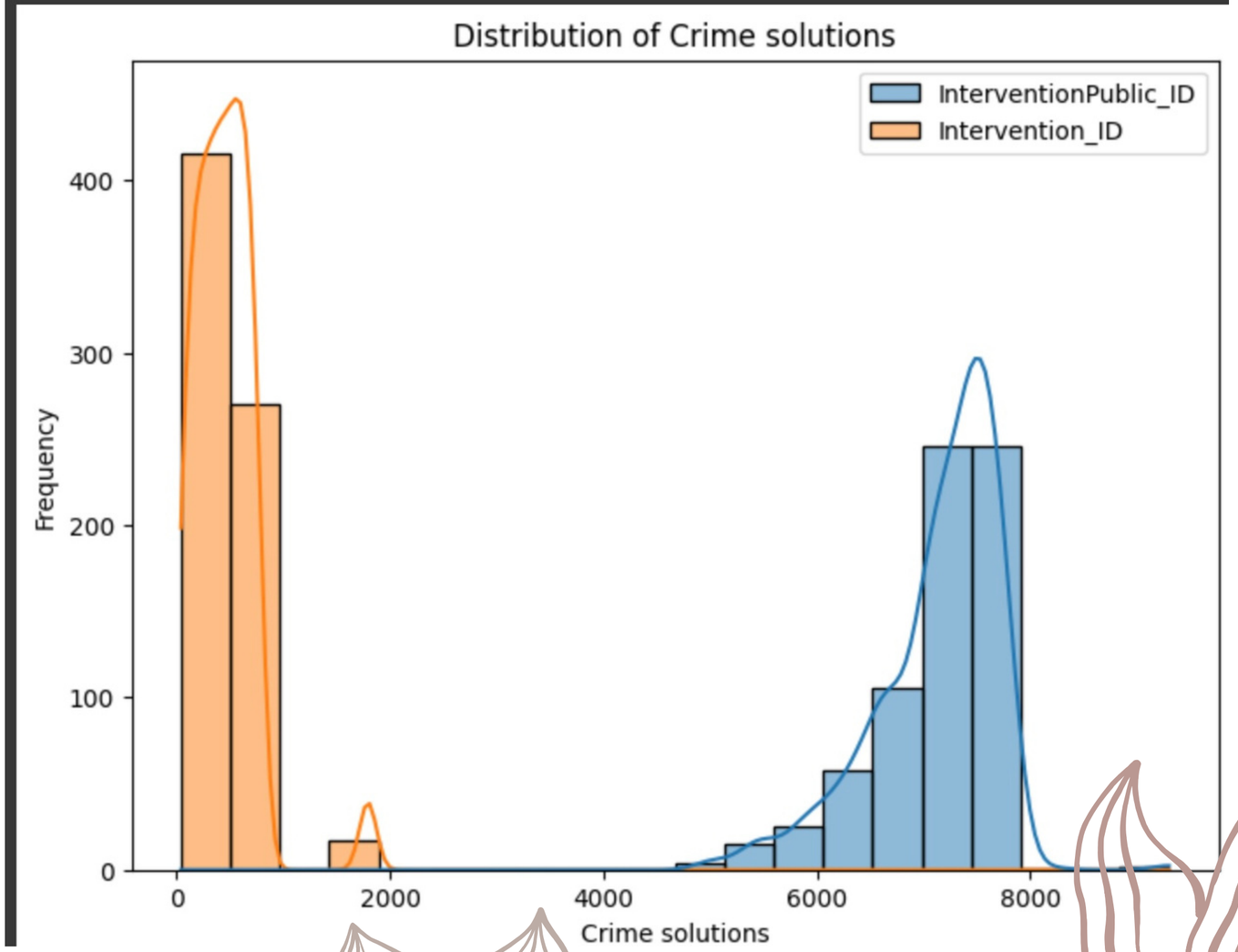
```
Summary statistics:
       InterventionPublic_ID  Intervention_ID
count             702.000000       702.000000
mean             7130.169516       452.699430
std               581.728225       298.067646
min              4995.000000        39.000000
25%              6843.250000       242.500000
50%              7270.000000       432.500000
75%              7549.500000       614.750000
max              9303.000000      1803.000000
```

```
# Unique values in categorical columns
print("\nUnique values in 'Geography' column:")
print(data['Geography'].unique())
```

```
Unique values in 'Geography' column:
['Urban' 'Suburban, Urban' 'Rural' nan 'Rural, Suburban, Urban'
 'Rural, Suburban' 'Rural, Tribal' 'Rural, Urban' 'Suburban'
 'Urban, Tribal' 'Rural, Suburban, Urban, Tribal' 'Tribal']
```

```
# Unique values in 'Ages' column
print("\nUnique values in 'Ages' column:")
print(data['Ages'].unique())
```

```
plt.figure(figsize=(8, 6))
sns.histplot(data, bins=20, kde=True)
plt.xlabel('Crime solutions')
plt.ylabel('Frequency')
plt.title('Distribution of Crime solutions')
plt.show()
```



SUMMARY + SOLUTIONS

Hypothesis or Question

1: THE GEOGRAPHIC LOCATION (E.G., 'URBAN,' 'SUBURBAN,' 'RURAL') OF CRIME SOLUTIONS IS ASSOCIATED WITH THEIR EFFECTIVENESS. URBAN PROGRAMS MAY HAVE DIFFERENT EFFECTIVENESS COMPARED TO SUBURBAN OR RURAL PROGRAMS.

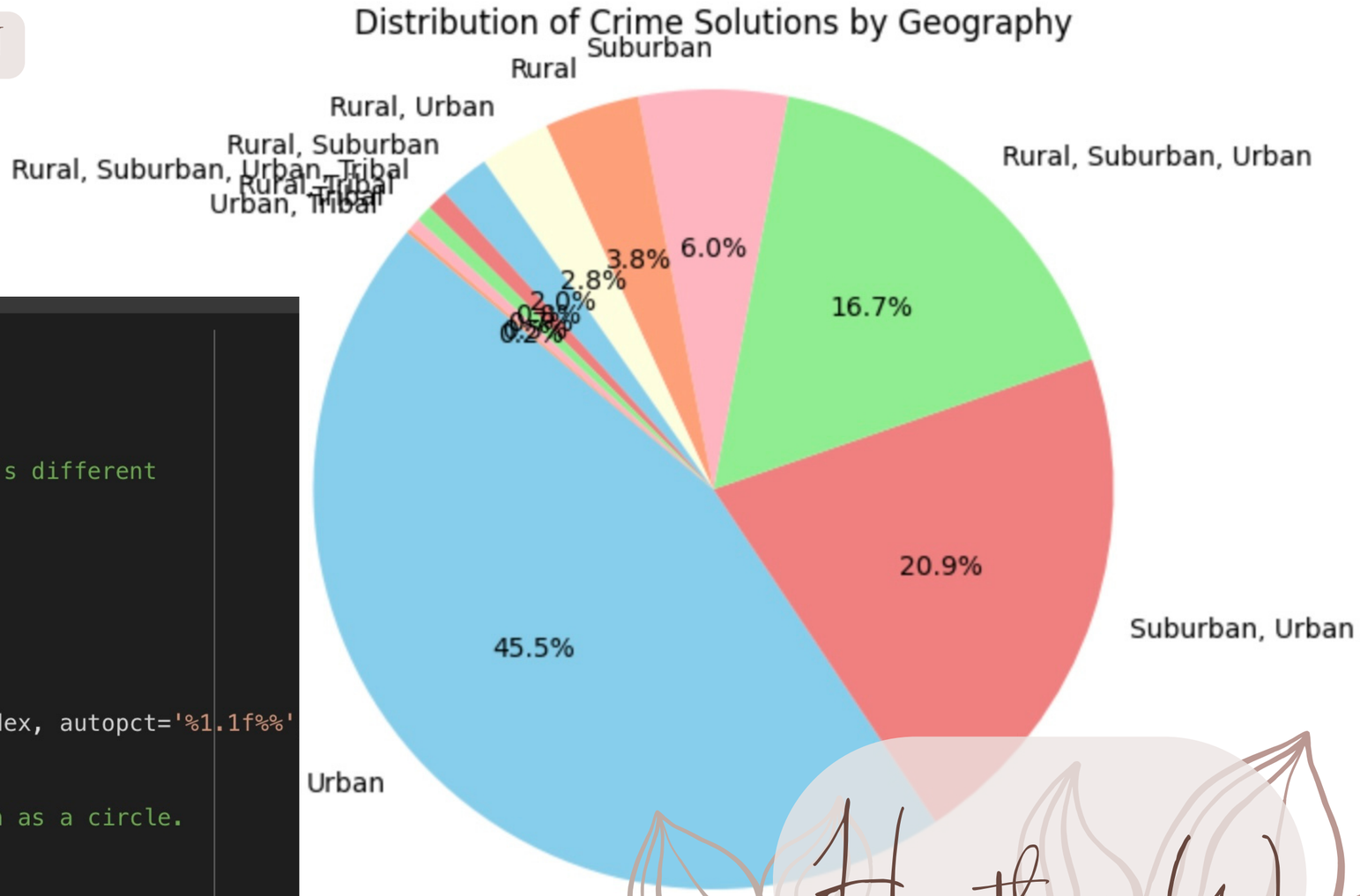
2: THE EVIDENCE RATING OF CRIME SOLUTIONS IS RELATED TO THEIR EFFECTIVENESS. SPECIFICALLY, PROGRAMS WITH A HIGHER EVIDENCE RATING WILL BE MORE EFFECTIVE IN REDUCING CRIME RATES.

3: CRIME SOLUTIONS TARGETING SPECIFIC GENDER DEMOGRAPHICS (E.G., 'MALE,' 'FEMALE,' 'BOTH') HAVE DIFFERENT EVIDENCE RATINGS. PROGRAMS DESIGNED FOR SPECIFIC GENDERS MAY HAVE DIFFERENT LEVELS OF EVIDENCE.

GEOGRAPHIC

THE GEOGRAPHIC LOCATION (E.G., 'URBAN,' 'SUBURBAN,' 'RURAL') OF CRIME SOLUTIONS IS ASSOCIATED WITH THEIR EFFECTIVENESS. URBAN PROGRAMS MAY HAVE DIFFERENT EFFECTIVENESS COMPARED TO SUBURBAN OR RURAL PROGRAMS.

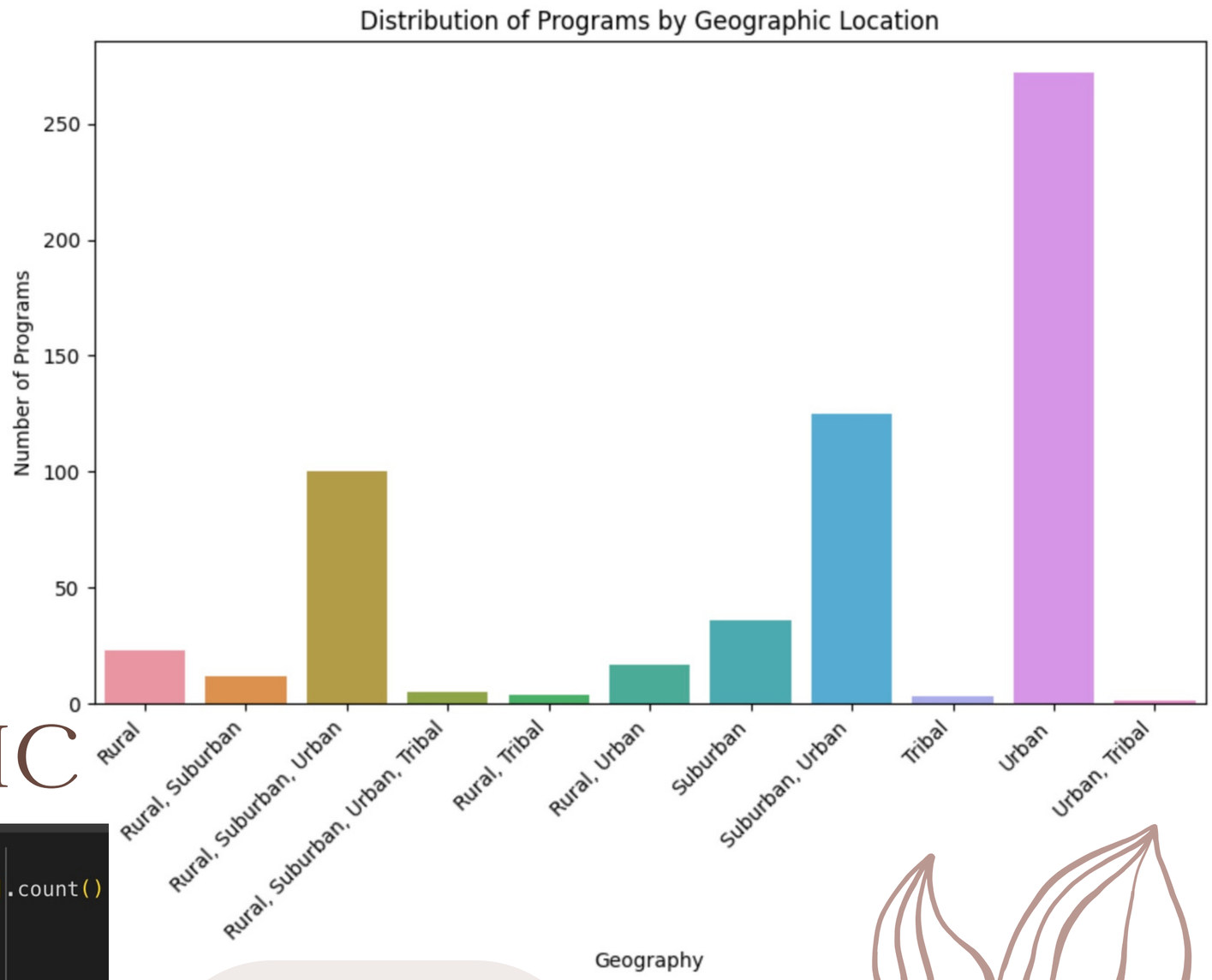
```python
import pandas as pd
import matplotlib.pyplot as plt

# Load your data (assuming you've already loaded the dataset)
# Replace 'data' with the variable containing your dataset if it's different
data = pd.read_csv('CrimeSolutions.gov_Programs.csv')

# Calculate the distribution of 'Geography'
geography_distribution = data['Geography'].value_counts()

# Create a pie chart to visualize the distribution
plt.figure(figsize=(10, 6))
plt.pie(geography_distribution, labels=geography_distribution.index, autopct='%1.1f%%'

# Customize the plot
plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.
plt.title('Distribution of Crime Solutions by Geography')

# Show the plot
plt.show()
```



Hypothesis(1.)

THE GEOGRAPHIC LOCATION (E.G., 'URBAN,' 'SUBURBAN,' 'RURAL') OF CRIME SOLUTIONS IS ASSOCIATED WITH THEIR EFFECTIVENESS. URBAN PROGRAMS MAY HAVE DIFFERENT EFFECTIVENESS COMPARED TO SUBURBAN OR RURAL PROGRAMS.

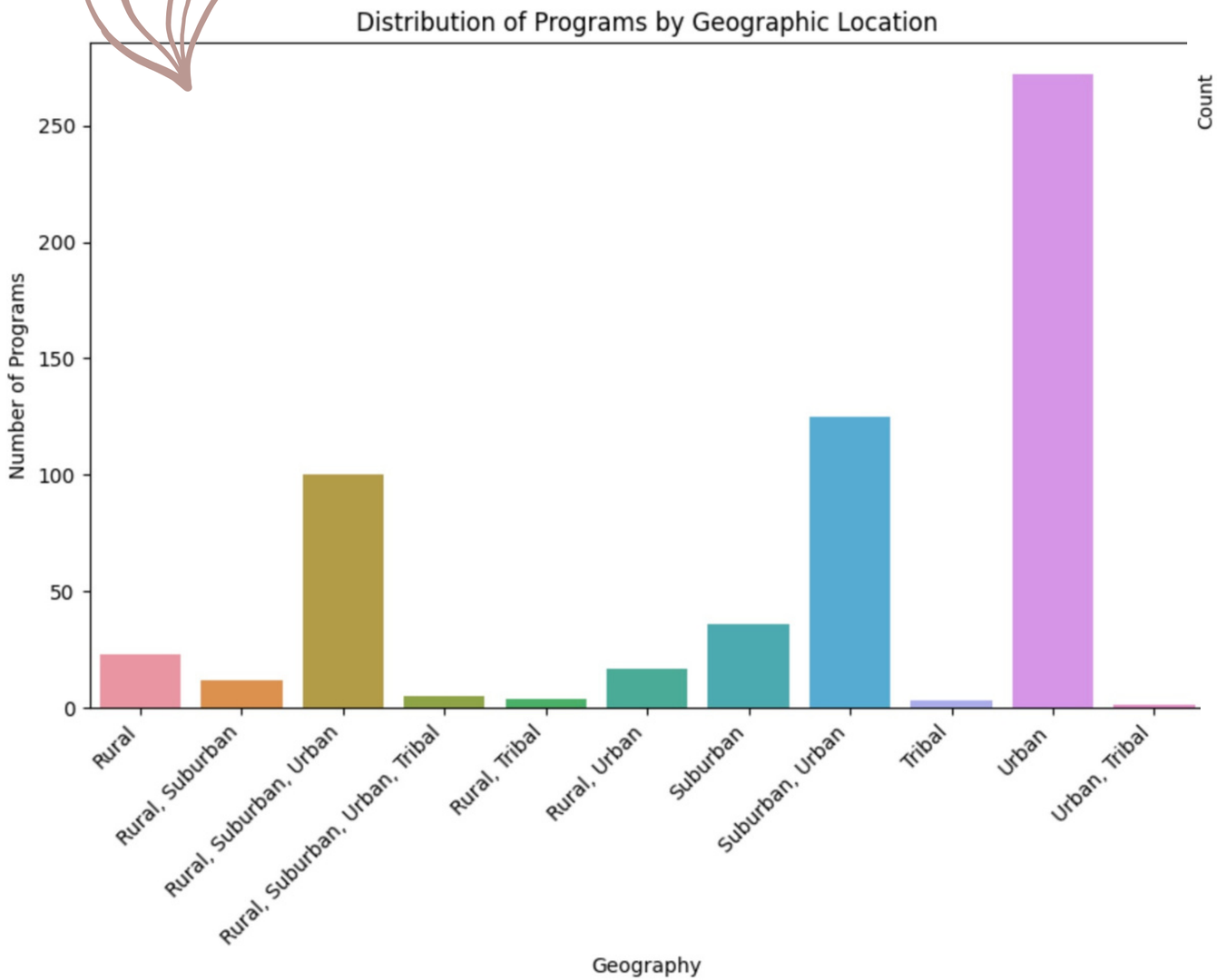
GEOGRAPHIC

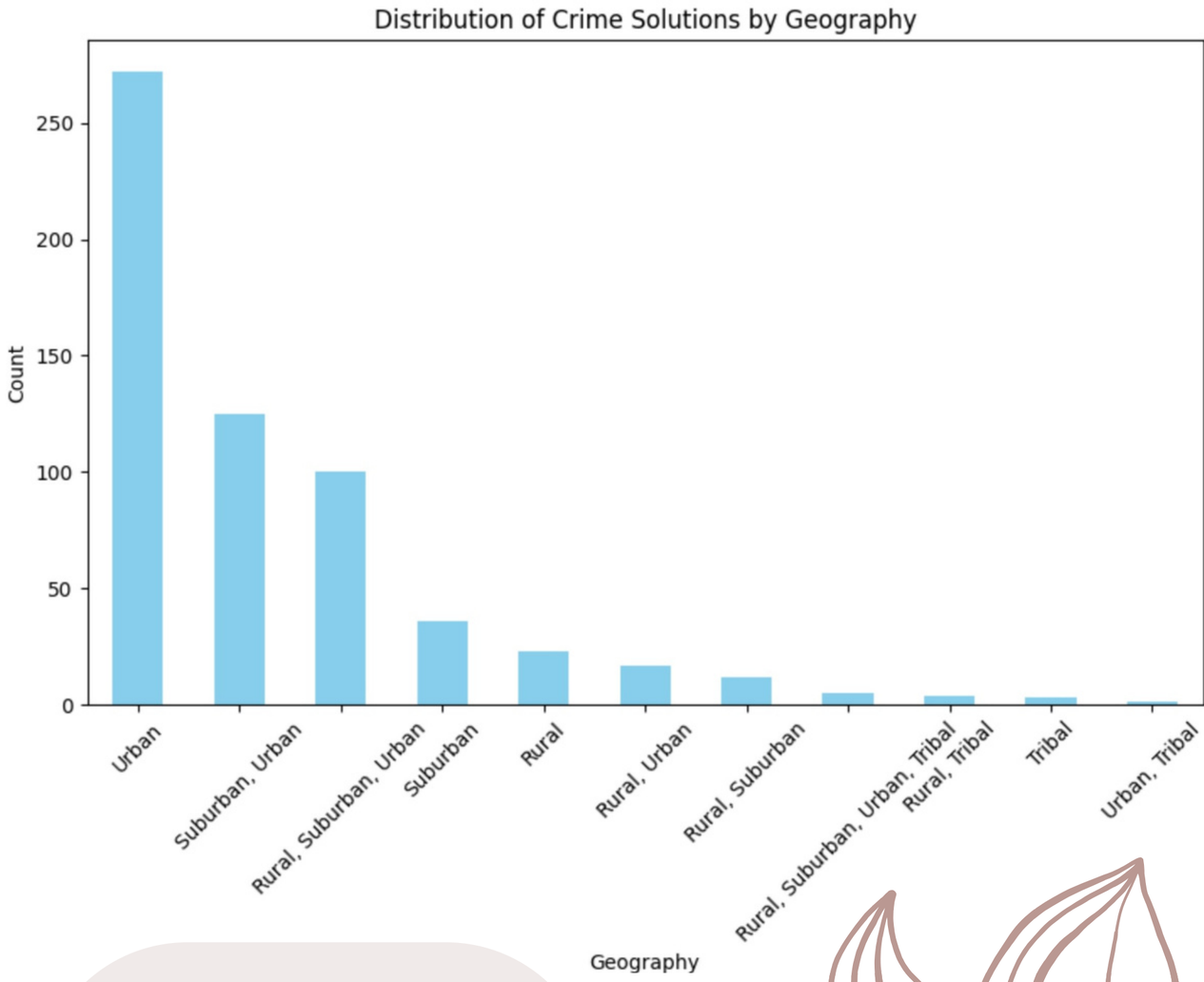



```python
# Group data by 'Geography' and count the number of programs in each location
program_counts_by_geography = data.groupby('Geography')['InterventionPublic_ID'].count()

# Visualize the distribution of programs by geographic location
plt.figure(figsize=(10, 6))
sns.barplot(x=program_counts_by_geography.index, y=program_counts_by_geography.values)
plt.xlabel('Geography')
plt.ylabel('Number of Programs')
plt.title('Distribution of Programs by Geographic Location')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Hypothesis(1.)

GEOGRAPHIC

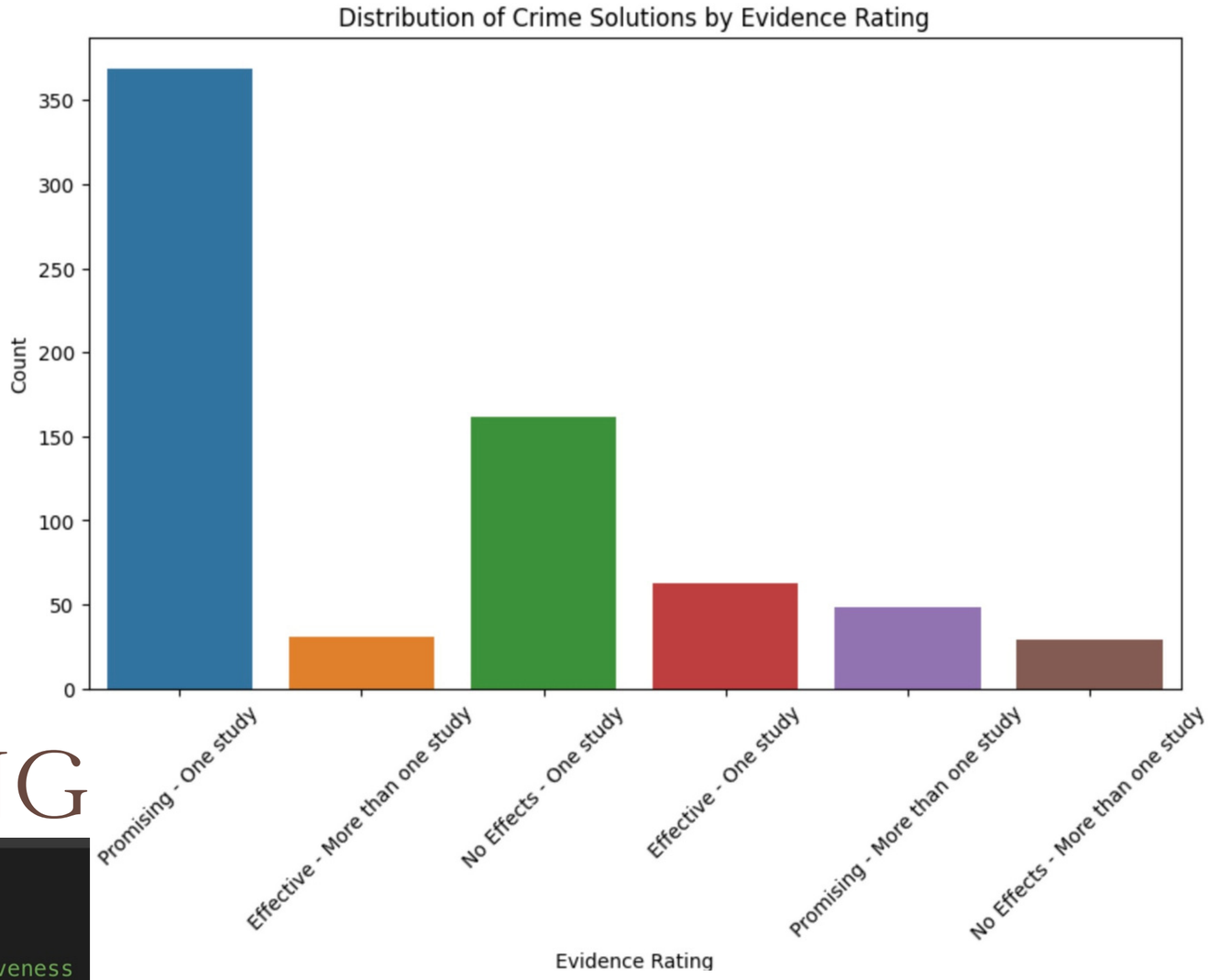




Hypothesis(1.)

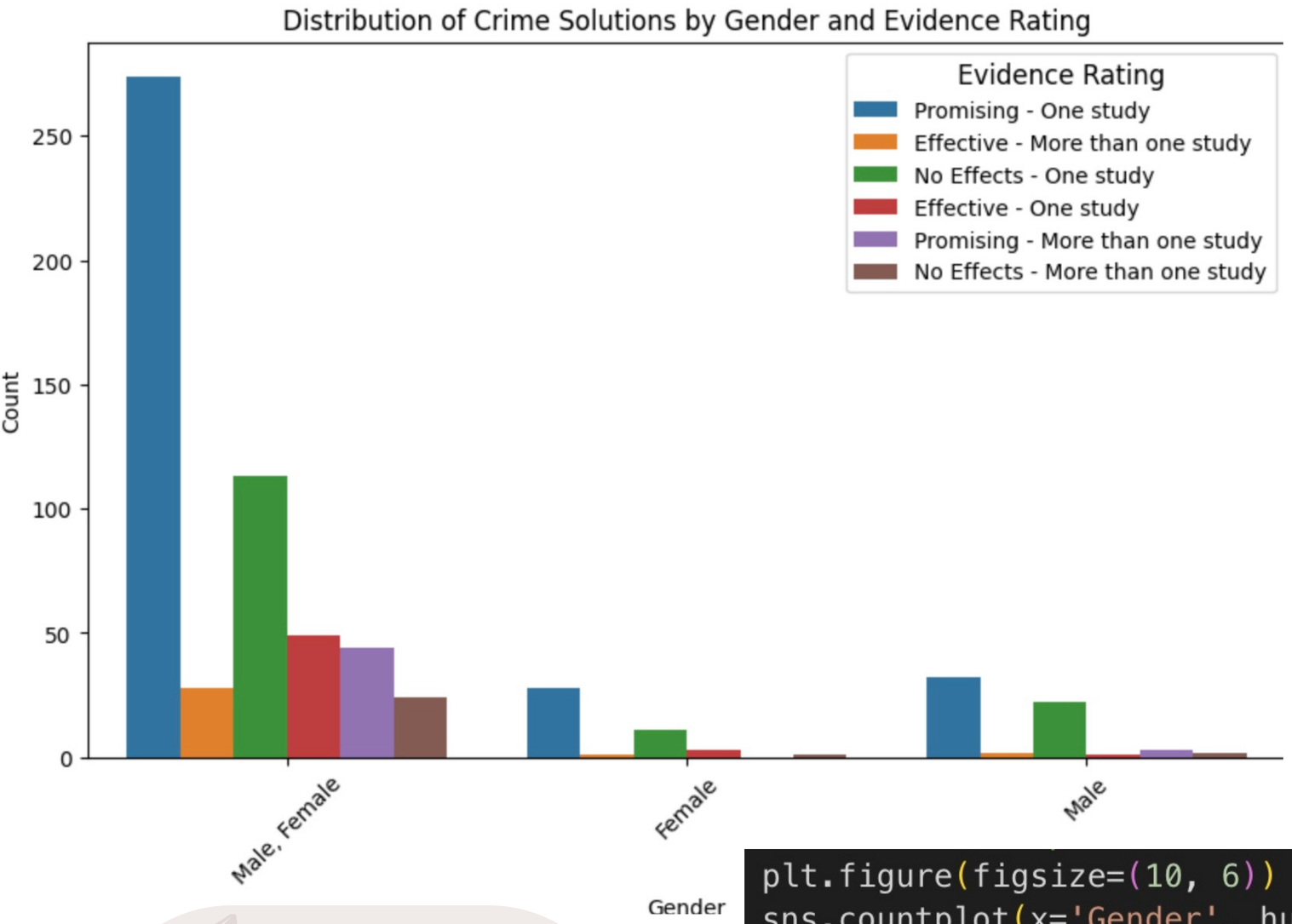
THE EVIDENCE RATING OF CRIME SOLUTIONS IS RELATED TO THEIR EFFECTIVENESS. SPECIFICALLY, PROGRAMS WITH A HIGHER EVIDENCE RATING WILL BE MORE EFFECTIVE IN REDUCING CRIME RATES.

EVIDENCE RATING




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

# Create a bar plot to visualize the relationship between Evidence Rating and Effectiveness
plt.figure(figsize=(10, 6))
sns.countplot(x='Evidence Rating', data=data)
plt.xlabel('Evidence Rating')
plt.ylabel('Count')
plt.title('Distribution of Crime Solutions by Evidence Rating')
plt.xticks(rotation=45)  # Rotate x-axis labels for better readability
plt.show()
```

Hypothesis(2.)



CRIME SOLUTIONS TARGETING SPECIFIC GENDER DEMOGRAPHICS (E.G., 'MALE,' 'FEMALE,' 'BOTH') HAVE DIFFERENT EVIDENCE RATINGS. PROGRAMS DESIGNED FOR SPECIFIC GENDERS MAY HAVE DIFFERENT LEVELS OF EVIDENCE.

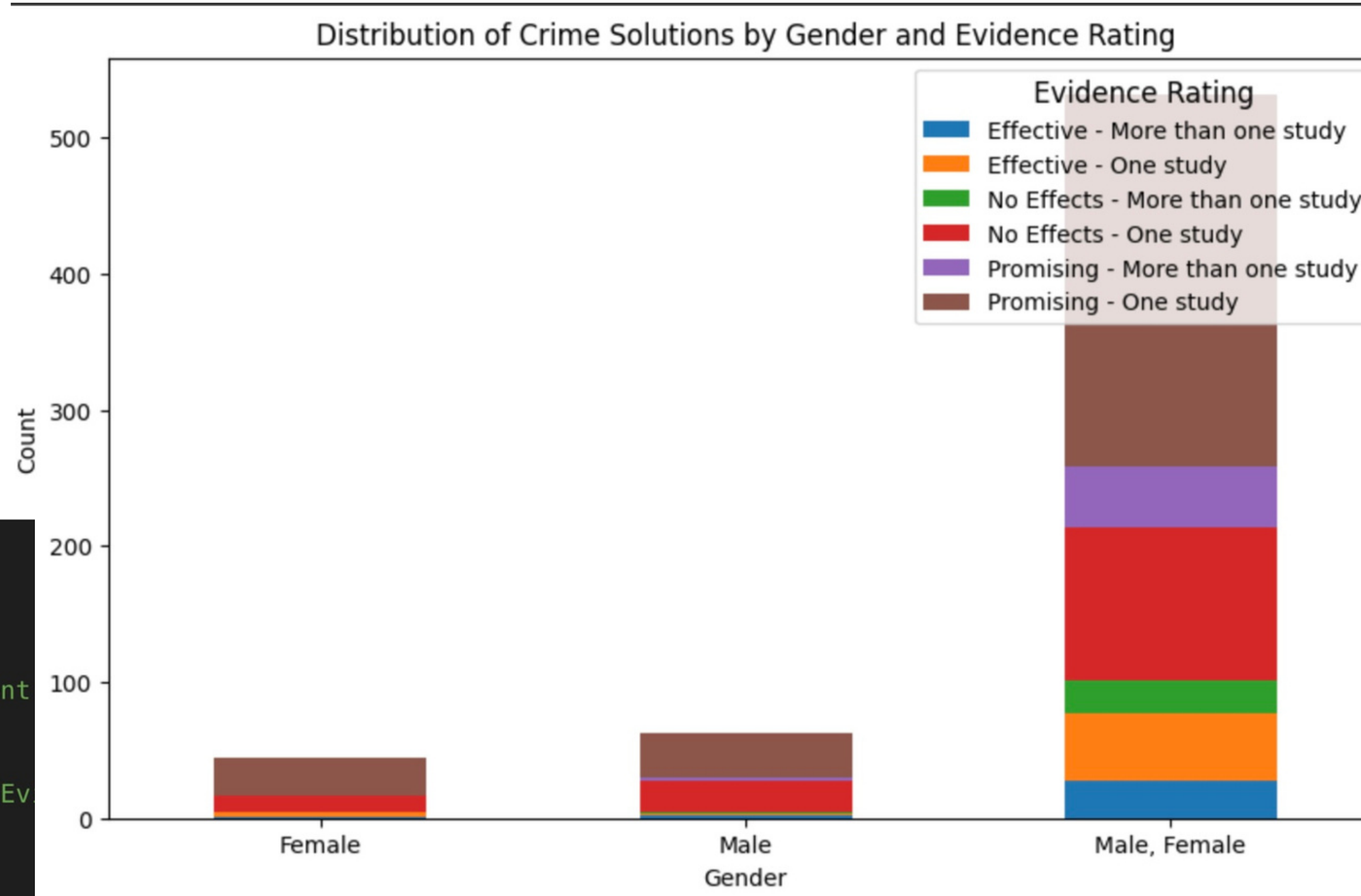
GENDER + EVIDENCE RATING

Hypothesis (3.)

```
plt.figure(figsize=(10, 6))
sns.countplot(x='Gender', hue='Evidence Rating', data=data)
plt.xlabel('Gender')
plt.ylabel('Count')
plt.title('Distribution of Crime Solutions by Gender and Evidence Rating')
plt.xticks(rotation=45)  # Rotate x-axis labels for better readability
plt.legend(title='Evidence Rating', title_fontsize='12', loc='upper right')
plt.show()
```

CRIME SOLUTIONS TARGETING SPECIFIC GENDER DEMOGRAPHICS (E.G., 'MALE,' 'FEMALE,' 'BOTH') HAVE DIFFERENT EVIDENCE RATINGS. PROGRAMS DESIGNED FOR SPECIFIC GENDERS MAY HAVE DIFFERENT LEVELS OF EVIDENCE.

```python
import matplotlib.pyplot as plt

# Load your data (assuming you've already loaded the dataset)
# Replace 'data' with the variable containing your dataset if it's different
data = pd.read_csv('CrimeSolutions.gov_Programs.csv')

# Create a pivot table to get counts of each combination of "Gender" and "Ev
pivot_table = pd.crosstab(data['Gender'], data['Evidence Rating'])

# Create a bar chart
pivot_table.plot(kind='bar', stacked=True, figsize=(10, 6))

# Customize the plot
plt.xlabel('Gender')
plt.ylabel('Count')
plt.title('Distribution of Crime Solutions by Gender and Evidence Rating')
plt.xticks(rotation=0)  # Set rotation angle for x-axis labels

# Display the legend
plt.legend(title='Evidence Rating', title_fontsize=12, loc='upper right')

# Show the plot
plt.show()
```

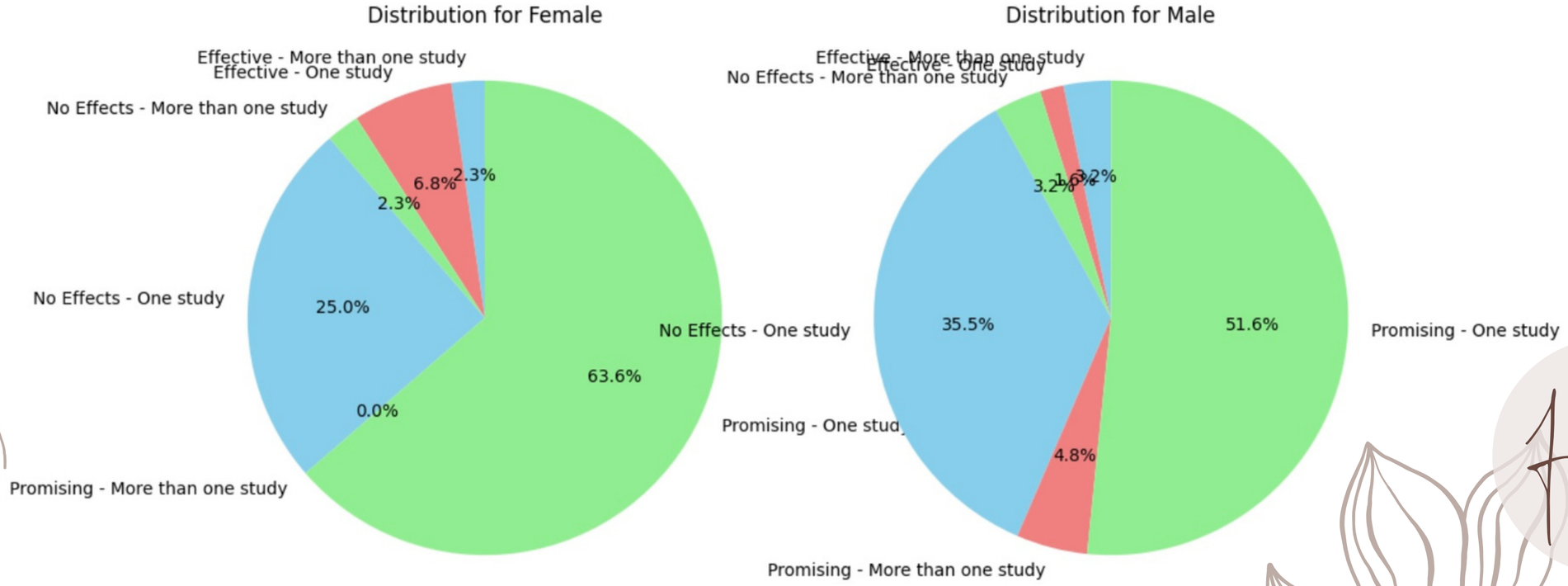


GENDER + EVIDENCE RATING

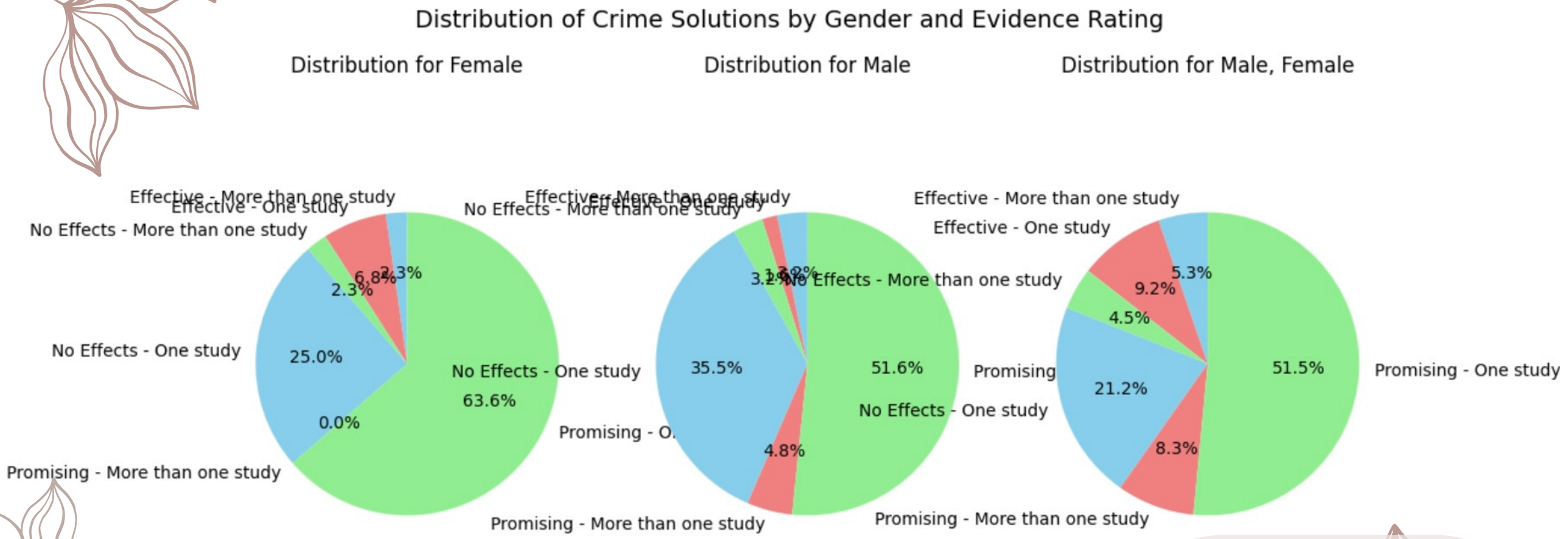
Hypothesis (3.)

GENDER + EVIDENCE RATING

```python
data = pd.read_csv('CrimeSolutions.gov_Programs.csv')

# Create a new DataFrame for the counts of each combination of "Gender" and "Evidence Rating"
gender_evidence_counts = data.groupby(['Gender', 'Evidence Rating']).size().unstack(fill_value=0)

# Create a pie chart for each gender category
fig, axes = plt.subplots(1, len(gender_evidence_counts), figsize=(12, 6))
colors = ['skyblue', 'lightcoral', 'lightgreen']

for i, gender in enumerate(gender_evidence_counts.index):
    ax = axes[i]
    ax.pie(gender_evidence_counts.loc[gender], labels=gender_evidence_counts.columns, colors=colors, autopct
    ax.set_title(f'Distribution for {gender}')
    ax.axis('equal')

plt.suptitle('Distribution of Crime Solutions by Gender and Evidence Rating', fontsize=14)
plt.show()
```



Hypothesis (3.)



GENDER + EVIDENCE RATING

Hypothesis (3.)