```
# NCRN Bird Monitoring - Project with Mock Data
# Created for college project using pandas, numpy, seaborn, matplotlib

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

# Step 1: Generate Mock Dataset
np.random.seed(42)  # For consistent output

parks = ['Rock Creek', 'Shenandoah', 'Prince William', 'Catoctin', 'Harpers Ferry']
years = np.random.choice(range(2007, 2018), size=1000)
species_codes = np.random.choice(['SP1', 'SP2', 'SP3', 'SP4', 'SP5', 'SP6', 'SP7', 'SP8',
'SP9', 'SP10'], size=1000)
points = np.random.randint(1, 100, size=1000)
conf_score = np.random.uniform(0.6, 1.0, size=1000)
parks_random = np.random.choice(parks, size=1000)

df = pd.DataFrame({
   'Year': years,
   'SpeciesCode': species_codes,
   'Point': points,
   'Park': parks_random,
   'ConfidenceScore': conf_score
})

# Step 2: Preview the Dataset
print(" Preview:\n", df.head())
print("\n Shape:", df.shape)
print("\n🕊️ Unique Species:", df['SpeciesCode'].nunique())

# Step 3: Top Species by Observation
top_species = df['SpeciesCode'].value_counts().head(10)
print("\n Top 10 Species:\n", top_species)

# Step 4: Observations by Year (Countplot)
plt.figure(figsize=(10, 5))
sns.countplot(data=df, x='Year', hue='Year', palette='crest', legend=False)
plt.title("Bird Observations per Year")
plt.xlabel("Year")
plt.ylabel("Observations")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# Step 5: Top Species Bar Chart (Barplot)
```

```
plt.figure(figsize=(10, 5))
sns.barplot(x=top_species.index, y=top_species.values, hue=top_species.index,
palette='flare', legend=False)
plt.title("Top 10 Bird Species")
plt.xlabel("Species Code")
plt.ylabel("Count")
plt.tight_layout()
plt.show()

# Step 6: Observations by Park (Countplot)
plt.figure(figsize=(10, 6))
sns.countplot(data=df, y='Park', hue='Park', order=df['Park'].value_counts().index,
palette='coolwarm', legend=False)
plt.title("Observations by Park")
plt.xlabel("Count")
plt.ylabel("Park")
plt.tight_layout()
plt.show()

# Step 7: Heatmap (Park vs Year)
pivot = df.pivot_table(index='Park', columns='Year', values='Point', aggfunc='count')
plt.figure(figsize=(12, 6))
sns.heatmap(pivot, annot=True, fmt='.0f', cmap='YlGnBu')
plt.title("Heatmap of Observations by Park & Year")
plt.xlabel("Year")
plt.ylabel("Park")
plt.tight_layout()
plt.show()

# Step 8: Confidence Score Distribution (Histogram)
plt.figure(figsize=(8, 4))
sns.histplot(df['ConfidenceScore'], kde=True, color='green', bins=20)
plt.title("Confidence Score Distribution (Generated with NumPy)")
plt.xlabel("Confidence Score")
plt.tight_layout()
plt.show()

# Step 9: Circle Plot - Scatter Plot
plt.figure(figsize=(8, 6))
sns.scatterplot(x='Year', y='Point', data=df, hue='SpeciesCode', size='ConfidenceScore',
sizes=(20, 200), palette='tab20', legend=False)
plt.title("Scatter Plot (Circle Plot)")
plt.xlabel("Year")
plt.ylabel("Point")
plt.tight_layout()
plt.show()
```

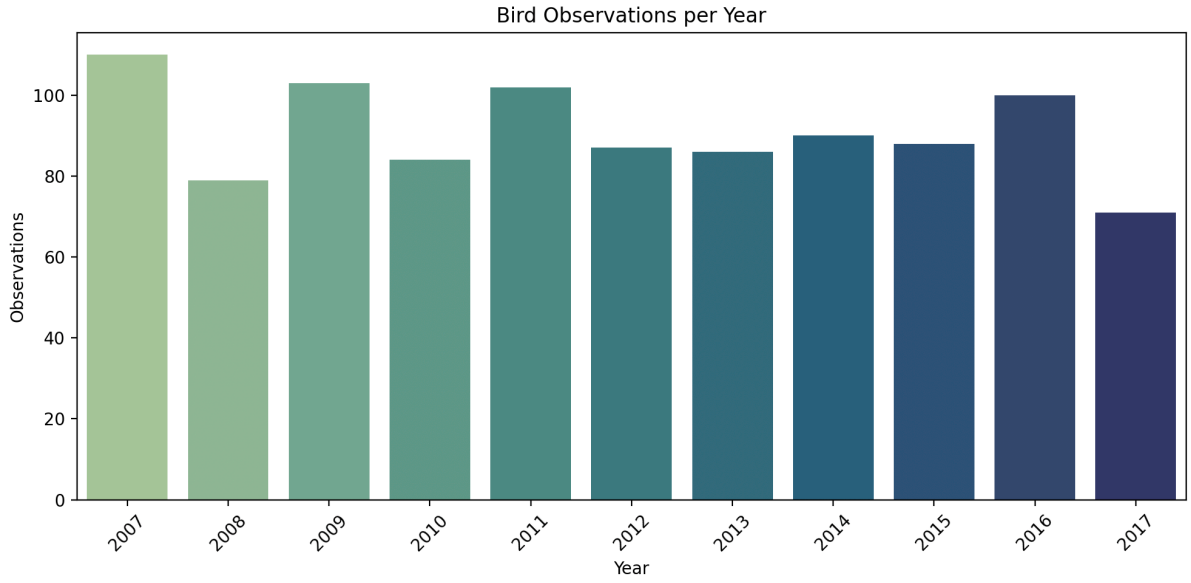
```
# Step 10: Circle Plot - Polar Plot (Radar Chart Example)
# Prepare data for polar plot (Example: Observations per Year)
year_counts = df['Year'].value_counts().sort_index()
angles = np.linspace(0, 2 * np.pi, len(year_counts), endpoint=False).tolist()
values = year_counts.values.tolist()

# Polar plot
plt.figure(figsize=(8, 8))
ax = plt.subplot(111, polar=True)
ax.fill(angles, values, color='orange', alpha=0.25)
ax.plot(angles, values, color='orange', linewidth=2)
ax.set_yticklabels([])
ax.set_xticks(angles)
ax.set_xticklabels(year_counts.index, fontsize=12)
plt.title("Polar Plot (Radar Chart) - Bird Observations by Year")
plt.tight_layout()
plt.show()

# Step 11: Boxplot for Confidence Score by Year
plt.figure(figsize=(10, 6))
sns.boxplot(data=df, x='Year', y='ConfidenceScore', hue='Year', palette='coolwarm',
legend=False)  # Fix: added hue
plt.title("Boxplot: Confidence Score by Year")
plt.xlabel("Year")
plt.ylabel("Confidence Score")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# Step 12: Pairplot for ConfidenceScore, Point, and Year
sns.pairplot(df[['ConfidenceScore', 'Point', 'Year']], hue='Year', palette='viridis')
plt.suptitle("Pairplot: Relationships Between Confidence Score, Point, and Year",
y=1.02)
plt.tight_layout()
plt.show()

# Step 13: Distribution of Points - KDE (Kernel Density Estimate)
plt.figure(figsize=(8, 6))
sns.kdeplot(df['Point'], fill=True, color='purple', alpha=0.5)  # Fix: changed shade to fill
plt.title("Density Plot: Point Distribution")
plt.xlabel("Point")
plt.ylabel("Density")
plt.tight_layout()
plt.show()

# Step 14: Correlation Heatmap (Confidence Score & Point)
correlation_matrix = df[['ConfidenceScore', 'Point']].corr()
```

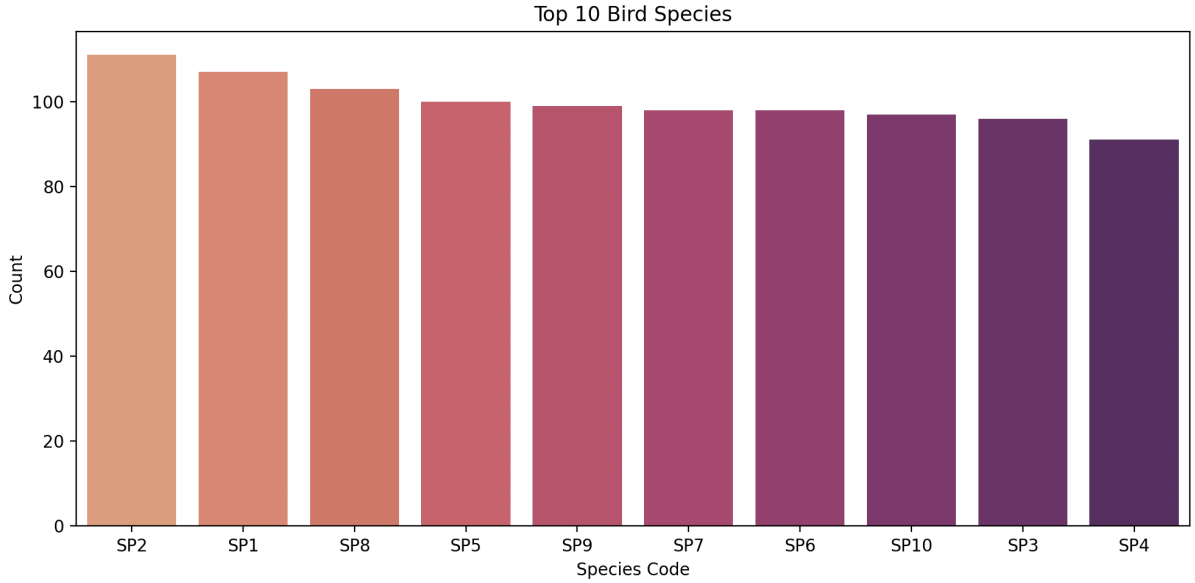
```
plt.figure(figsize=(6, 4))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f', vmin=-1,
vmax=1)
plt.title("Correlation Heatmap: ConfidenceScore & Point")
plt.tight_layout()
plt.show()

# Step 15: Pie Chart for Species Distribution
species_count = df['SpeciesCode'].value_counts()
plt.figure(figsize=(8, 8))
species_count.plot.pie(autopct='%1.1f%%', colors=sns.color_palette("Set3",
len(species_count)), startangle=90, legend=False)
plt.title("Pie Chart: Distribution of Bird Species")
plt.ylabel("")  # Remove ylabel for aesthetic purposes
plt.tight_layout()
plt.show()

# Step 16: Top 5 Parks with Most Observations (Bar Plot)
top_parks = df['Park'].value_counts().head(5)
plt.figure(figsize=(10, 6))
sns.barplot(x=top_parks.index, y=top_parks.values, hue=top_parks.index,
palette='coolwarm', legend=False)  # Fix: added hue
plt.title("Top 5 Parks with Most Observations")
plt.xlabel("Park")
plt.ylabel("Number of Observations")
plt.tight_layout()
plt.show()

# Step 17: Bird Observations Over Time (Time Series Plot)
observations_by_year = df.groupby('Year').size()
plt.figure(figsize=(10, 5))
observations_by_year.plot(kind='line', marker='o', color='darkorange')
plt.title("Bird Observations Over Time")
plt.xlabel("Year")
plt.ylabel("Number of Observations")
plt.tight_layout()
plt.show()

# Step 18: Distribution of Observations by Park (Violin Plot)
plt.figure(figsize=(10, 6))
sns.violinplot(data=df, x='Park', y='Point', hue='Park', palette='muted', legend=False)  #
Fix: added hue
plt.title("Distribution of Observations by Park")
plt.xlabel("Park")
plt.ylabel("Number of Observations")
plt.xticks(rotation=45)
plt.tight_layout()
```

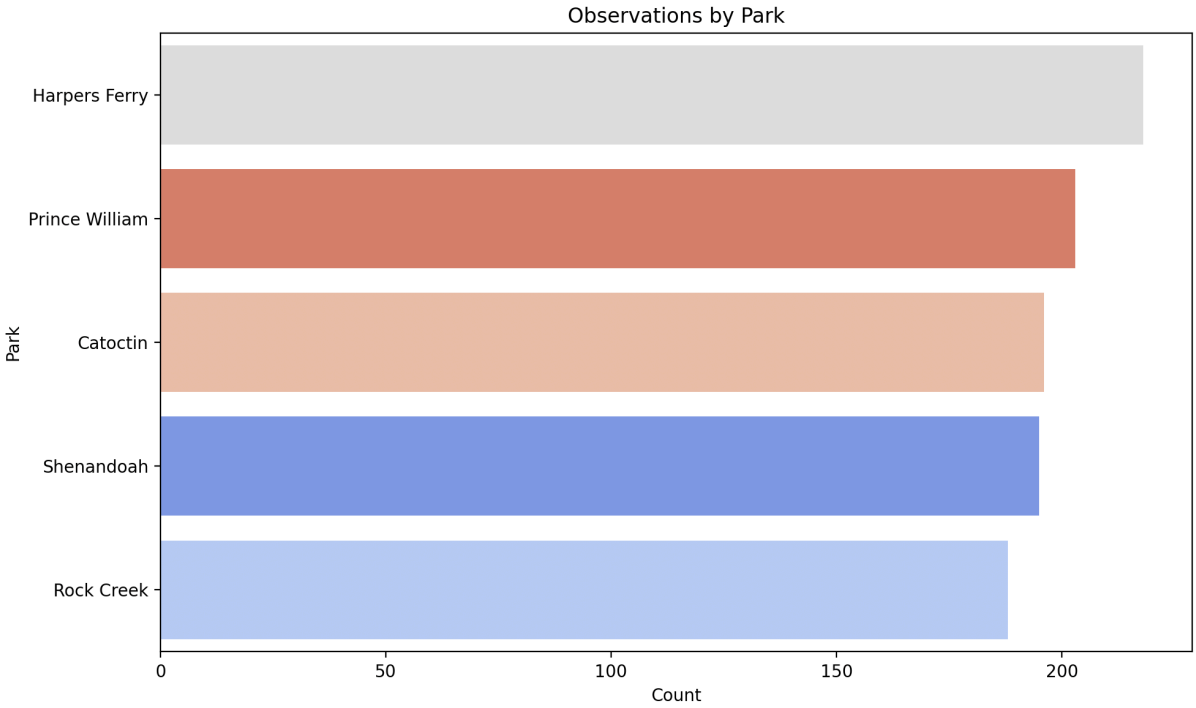
```
plt.show()

# Step 19: Top Species by Park (Stacked Bar Plot)
top_species_by_park = df.groupby(['Park', 'SpeciesCode']).size().unstack().fillna(0)
top_species_by_park = top_species_by_park.T  # Transpose to make species columns

plt.figure(figsize=(12, 8))
top_species_by_park.plot(kind='bar', stacked=True, colormap='Set3', figsize=(12, 6))
plt.title("Top Species by Park (Stacked Bar Plot)")
plt.xlabel("Species Code")
plt.ylabel("Observations")
plt.tight_layout()
plt.show()

# Final Message
print("\n Mock Data Project Completed Successfully!")
```

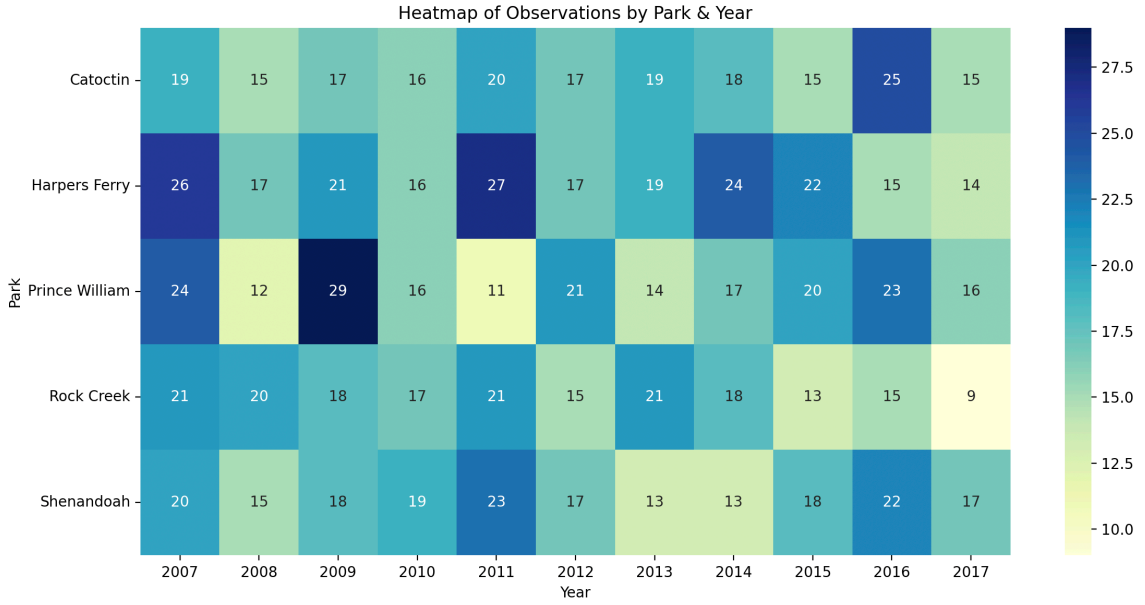
```
Preview:
   Year SpeciesCode  Point         Park  ConfidenceScore
0  2013        SP7    39    Shenandoah         0.903699
1  2010        SP8    13    Rock Creek         0.961443
2  2017        SP3    31  Harpers Ferry        0.693105
3  2014        SP6    71  Harpers Ferry        0.922958
4  2011        SP1    63    Shenandoah         0.623638

 Shape: (1000, 5)

🕊️ Unique Species: 10

 Top 10 Species:
 SpeciesCode
SP2    111
SP1    107
SP8    103
SP5    100
SP9     99
```

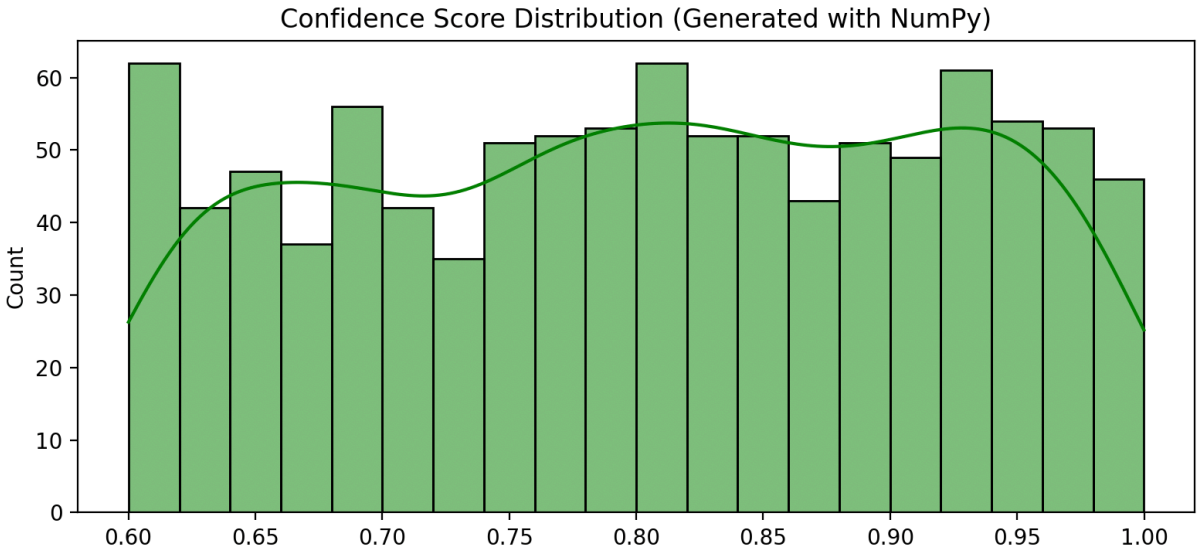
SP7 98
SP6 98
SP10 97
SP3 96
SP4 91
Name: count, dtype: int64

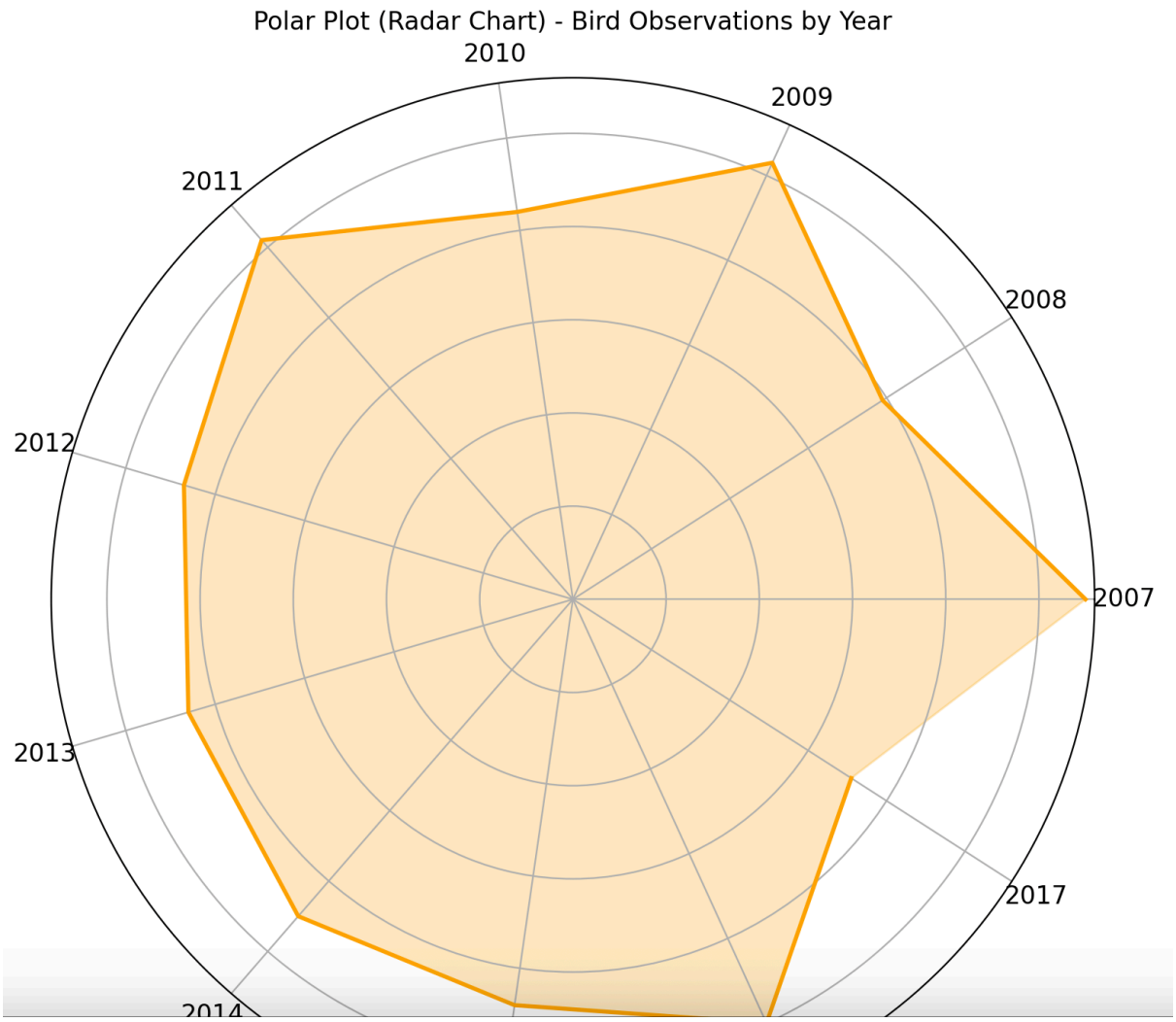


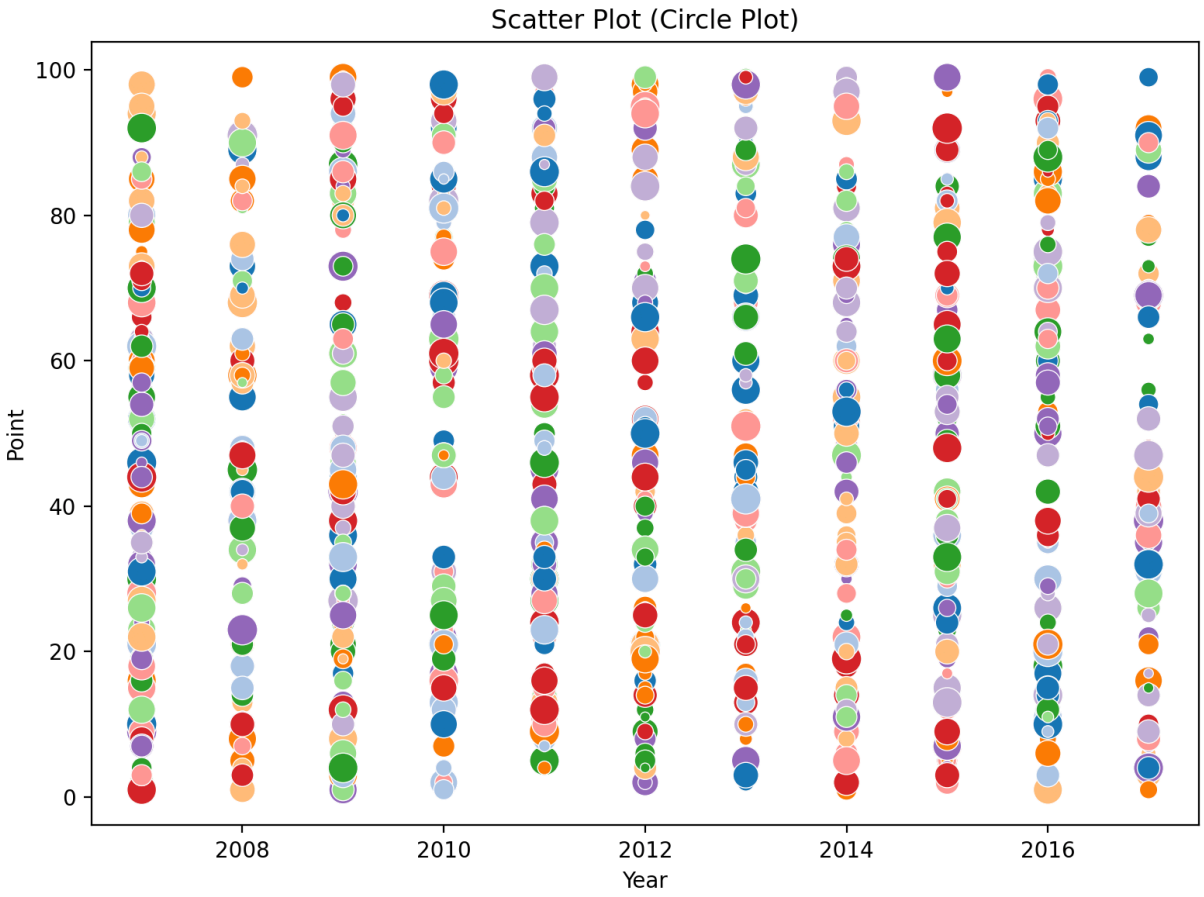


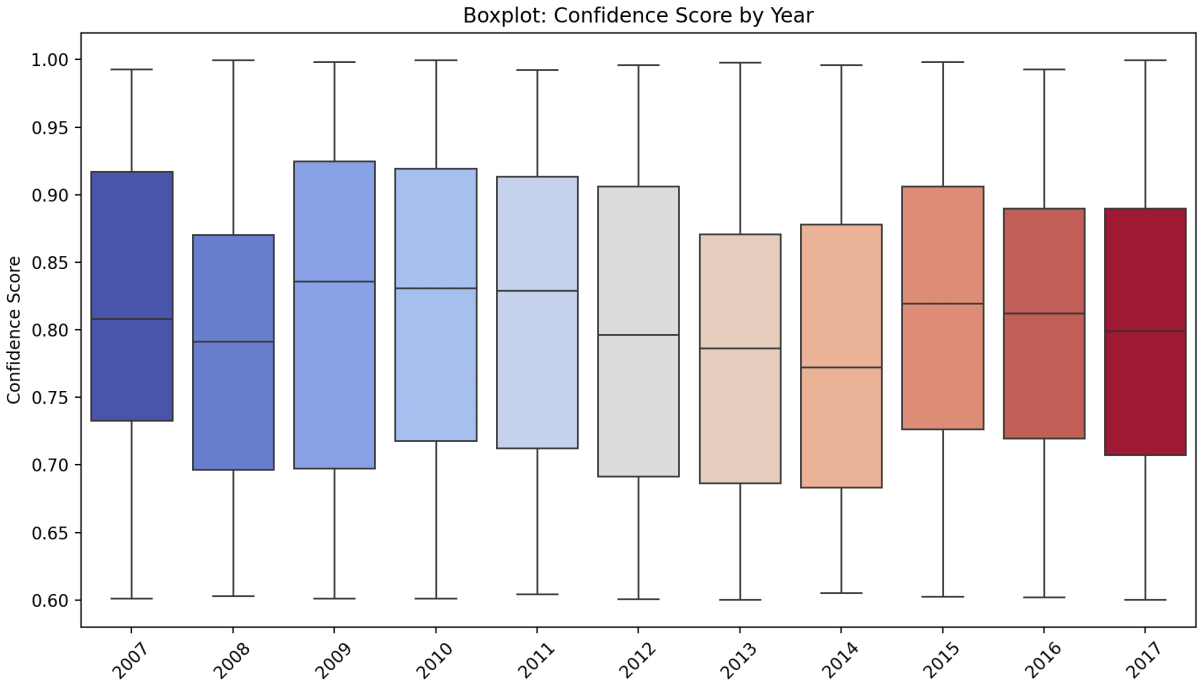

Observations by Park
Harpers Ferry
Prince William
Catoctin
Shenandoah
Rock Creek
Park
0
50
100
150
200
Count

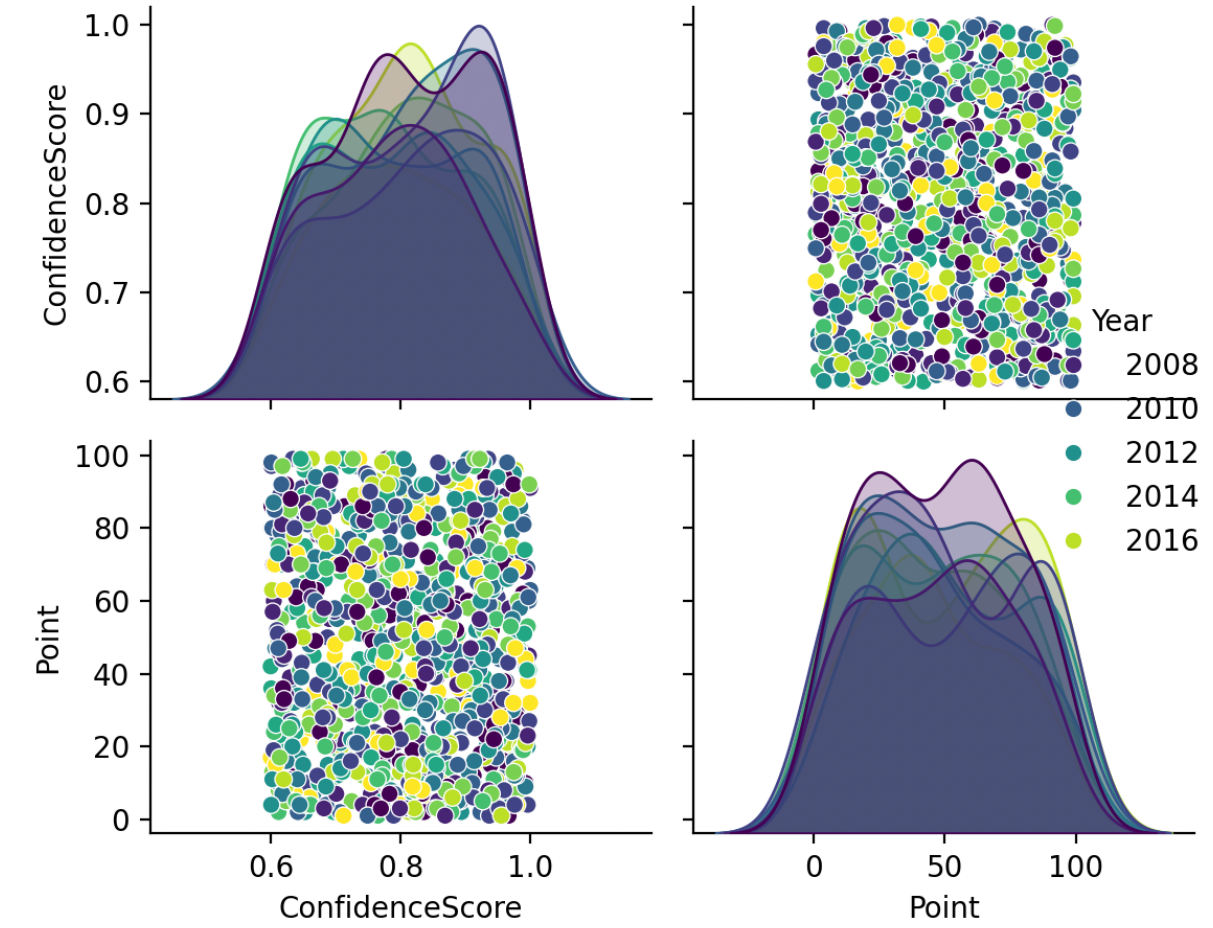

Heatmap of Observations by Park & Year
Catoctin 19 15 17 16 20 17 19 18 15 25 15
Harpers Ferry 26 17 21 16 27 17 19 24 22 15 14
Prince William 24 12 29 16 11 21 14 17 20 23 16
Rock Creek 21 20 18 17 21 15 21 18 13 15 9
Shenandoah 20 15 18 19 23 17 13 13 18 22 17
Park
2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017
Year
27.5
25.0
22.5
20.0
17.5
15.0
12.5
10.0

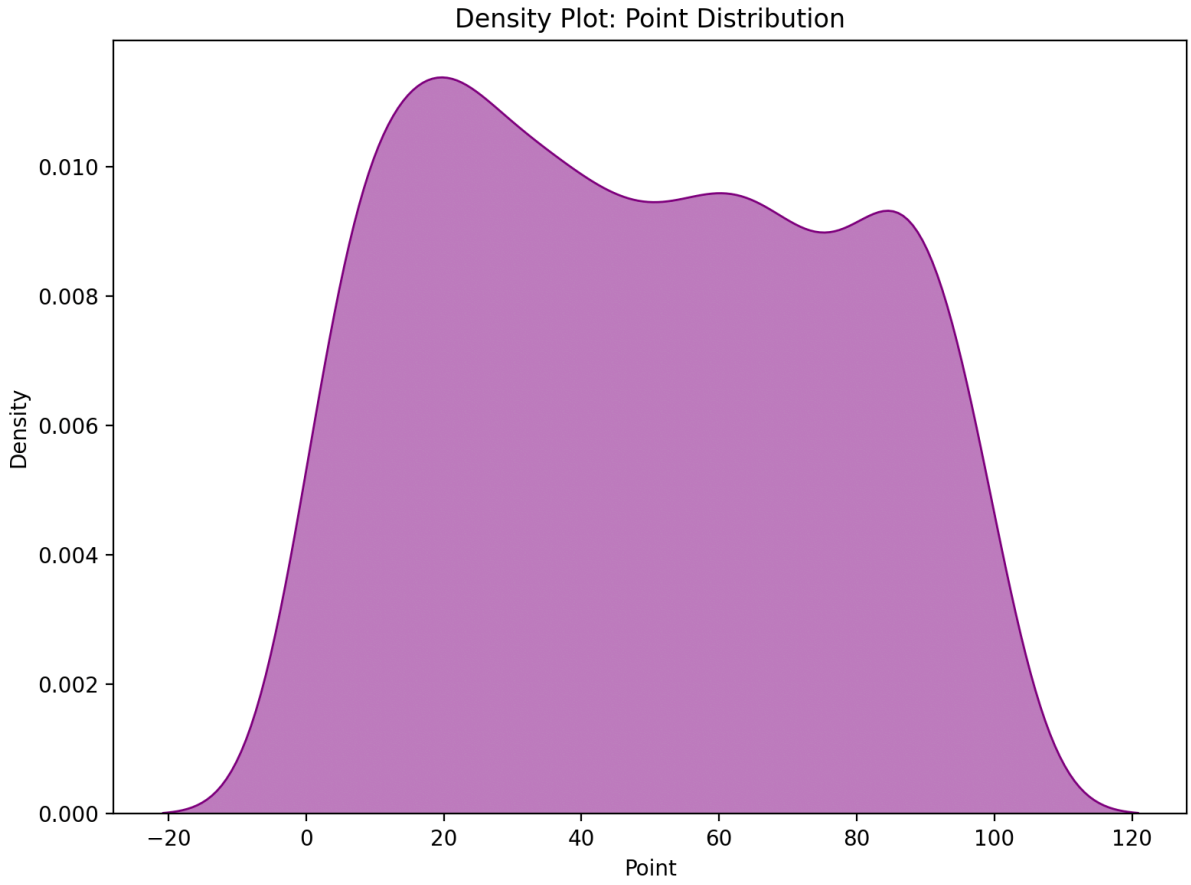

Confidence Score Distribution (Generated with NumPy)
Count
0
10
20
30
40
50
60
0.60
0.65
0.70
0.75
0.80
0.85
0.90
0.95
1.00

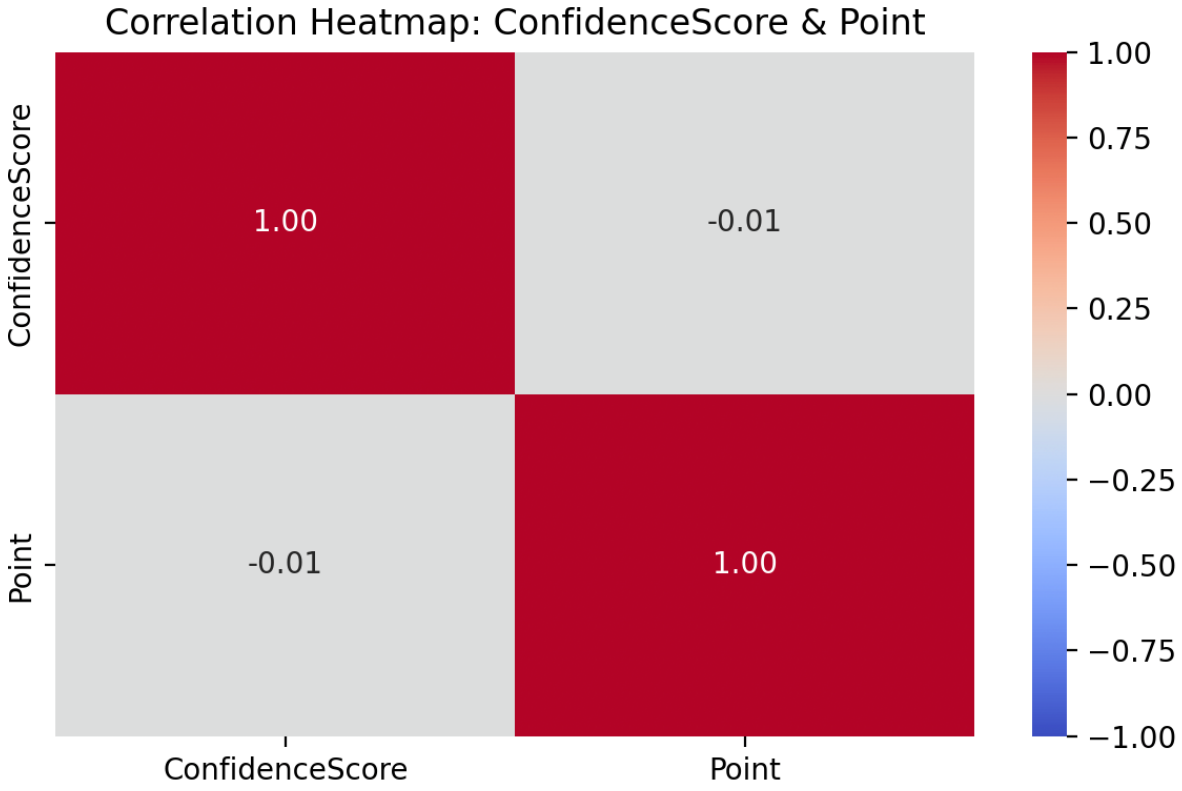

Polar Plot (Radar Chart) - Bird Observations by Year
2010
2009
2008
2007
2017
2014
2013
2012
2011


Scatter Plot (Circle Plot)
Point
Year
0
20
40
60
80
100
2008
2010
2012
2014
2016

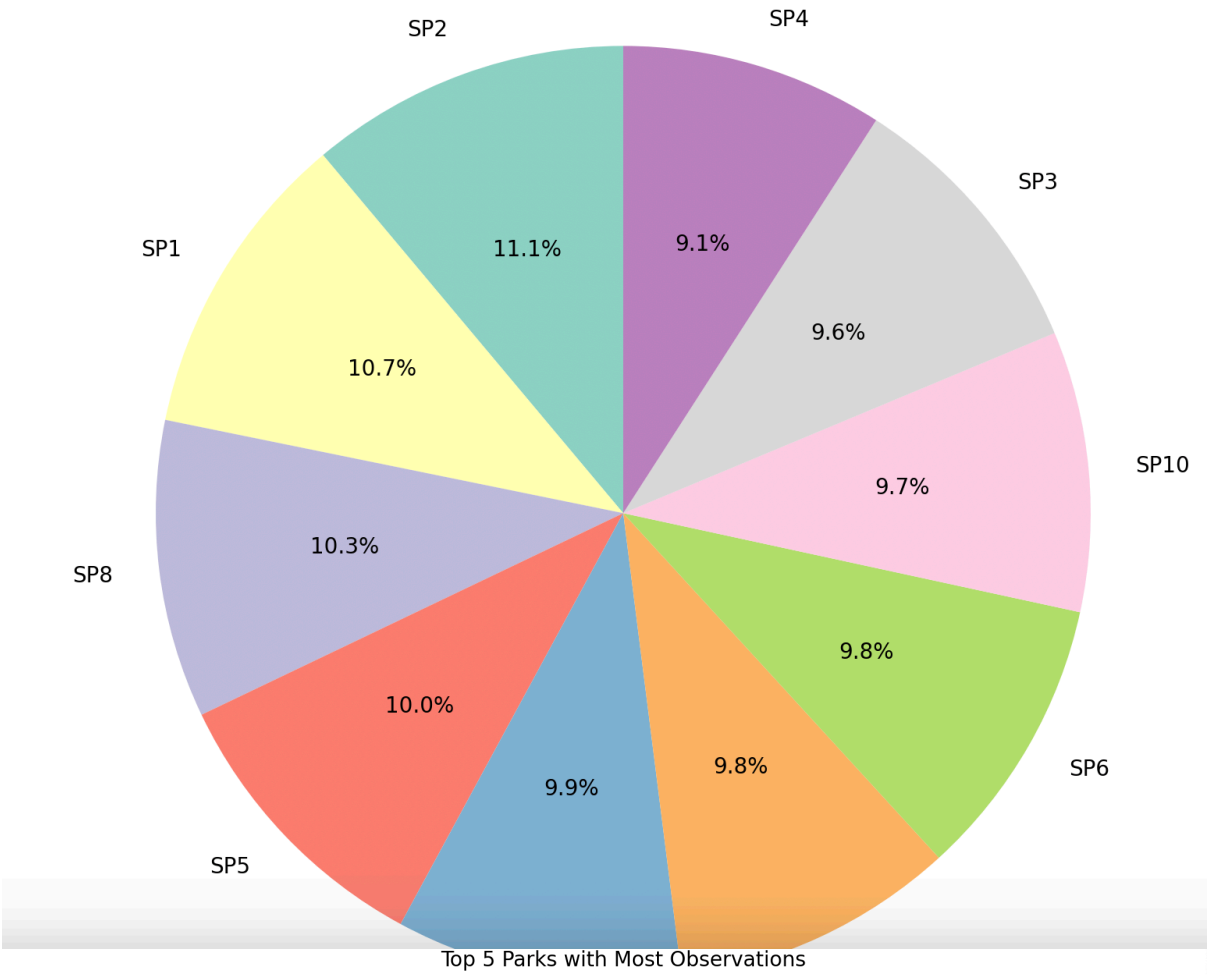

Boxplot: Confidence Score by Year
Confidence Score
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017


ConfidenceScore
1.0
0.9
0.8
0.7
0.6
Point
100
80
60
40
20
0
0.6
0.8
1.0
ConfidenceScore
0
50
100
Point
Year
2008
2010
2012
2014
2016

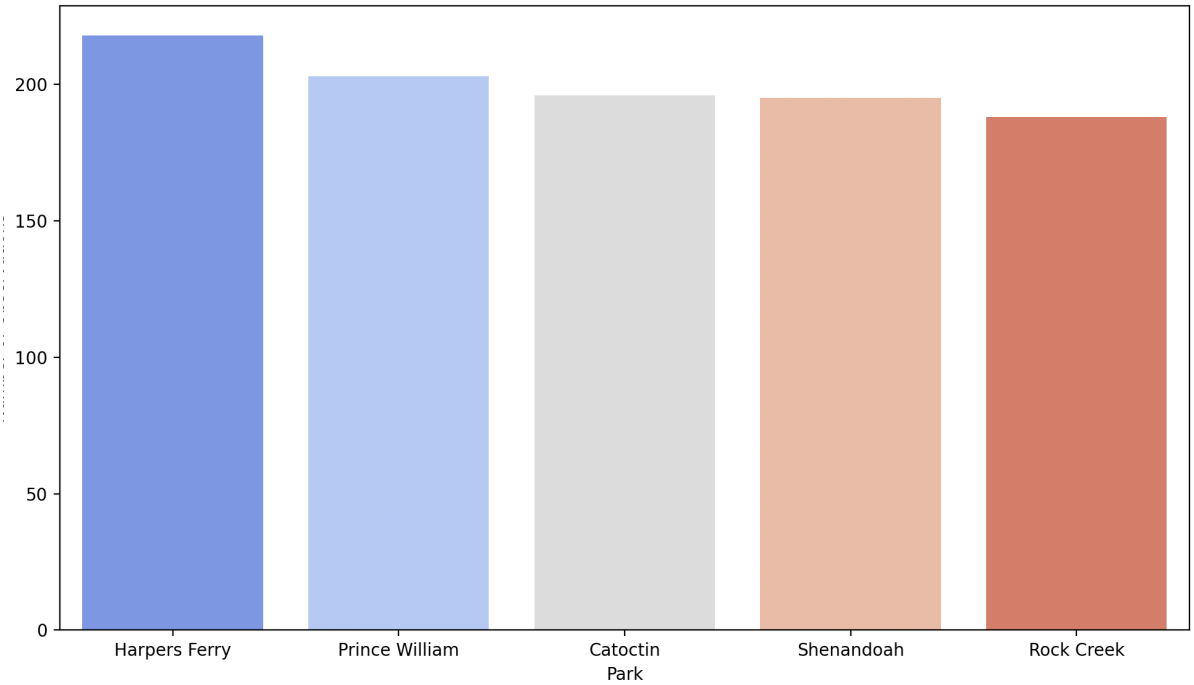

Density Plot: Point Distribution
Density
0.010
0.008
0.006
0.004
0.002
0.000
−20
0
20
40
60
80
100
120
Point

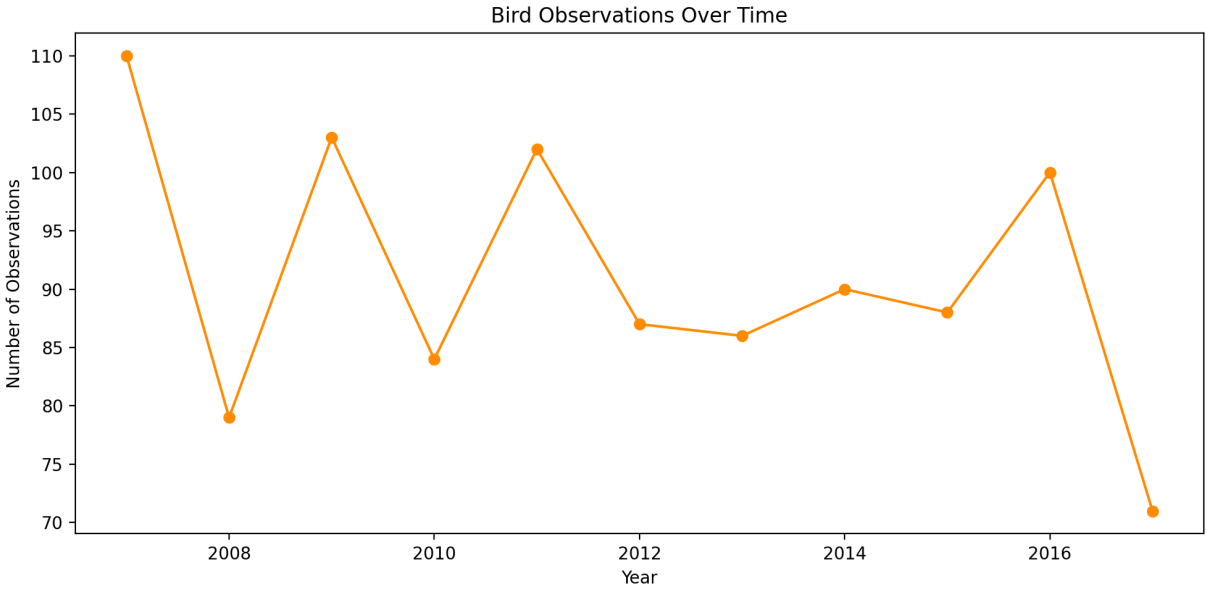

Correlation Heatmap: ConfidenceScore & Point
ConfidenceScore
Point
1.00
-0.01
-0.01
1.00
ConfidenceScore
Point
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

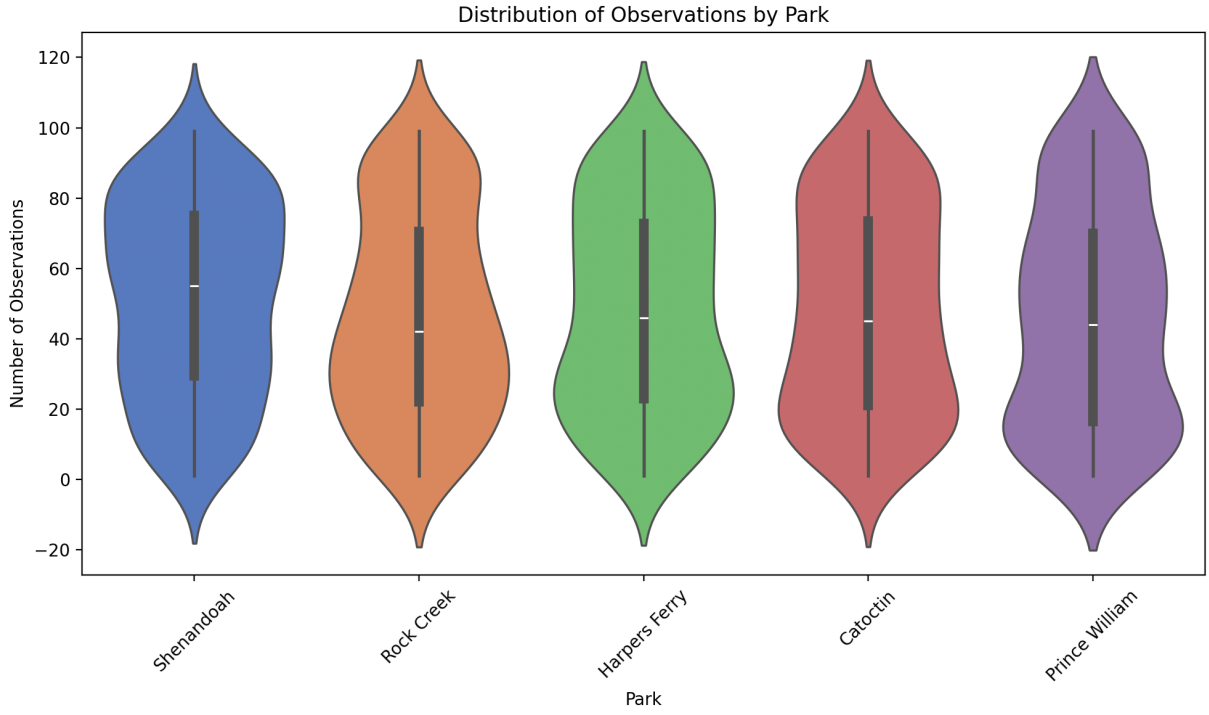
Pie Chart: Distribution of Bird Species

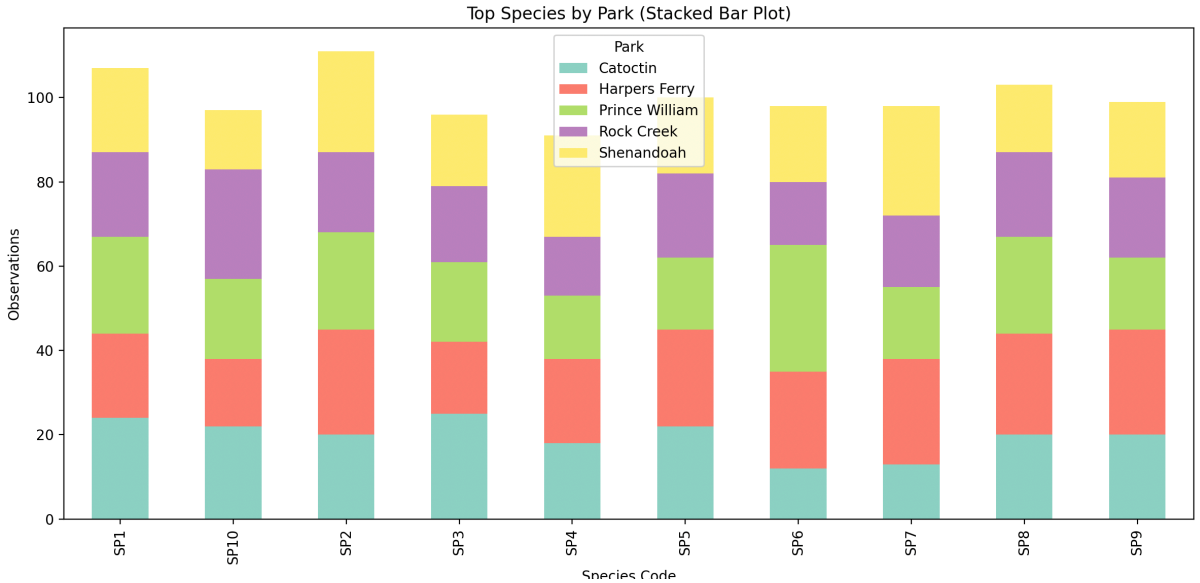


Top 5 Parks with Most Observations




Bird Observations Over Time
Number of Observations
110
105
100
95
90
85
80
75
70
2008
2010
2012
2014
2016
Year


Distribution of Observations by Park
Number of Observations
120
100
80
60
40
20
0
−20
Shenandoah
Rock Creek
Harpers Ferry
Catoctin
Prince William
Park


Top Species by Park (Stacked Bar Plot)
Park
Catoctin
Harpers Ferry
Prince William
Rock Creek
Shenandoah
Observations
0
20
40
60
80
100
SP1
SP10
SP2
SP3
SP4
SP5
SP6
SP7
SP8
SP9
Species Code