TOKYO INTERNATIONAL UNIVERSITY

PYTHON FOR DATA SCIENCE

TREND ANALYSIS OF HATE CRIMES IN THE US

27 OCTOBER, 2026

Lancelot Tabiaat - Adiya Koblandy - Toan

ANALYSIS OF HATE CRIMES IN THE US

INTRODUCTION

Hate crime ***Data Investigation*** using the FBI crime data explorer (CDE) API.

Goal: Pattern and Relationship ***Identification*** between *hate crime types, offender demographics, and time trends.*

DATA COLLECTION & CLEANING

Necessary Libraries

```python
import os, math, time, requests
import pandas as pd, numpy as np
from datetime import datetime
import matplotlib.pyplot as plt
import json as js
import scipy.stats as stats
import pprint as pprint
```

Data to collect

```python
headers = {
    'accept': 'application/json',
}

df_location_type = pd.DataFrame()
df_offender_race = pd.DataFrame()
df_offense_type = pd.DataFrame()
df_victim_type = pd.DataFrame()
df_bias = pd.DataFrame()
df_bias_category = pd.DataFrame()
df_monthly_totals = pd.DataFrame()
```

Automated Data Collection Loop (1995-2024)

```python
for year in years:
    for i, month in enumerate(months):
        if i == 11:
            month_next = '01'
            year_next = year + 1
        else:
            month_next = months[i + 1]
            year_next = year
```

```python
params = {
    'from': f'{month}-{year}',
    'to': f'{month_next}-{year_next}',
    'API_KEY': 'iiHnOKfno2Mgkt5AynpvPpUQTEyxE77jo1RU8PIv',
}

response1 = requests.get(
    'https://api.usa.gov/crime/fbi/cde/hate-crime/national?type=totals',
    params=params,
    headers=headers
)

response2 = requests.get(
    'https://api.usa.gov/crime/fbi/cde/hate-crime/national?type=counts',
    params=params,
    headers=headers
)
```

DATA COLLECTION & CLEANING

Top 7 Hate Crime Categories by Frequency

```
top7_cols_location = df_location_type.sum().sort_values(ascending=False).head(7).index
df_location_type = df_location_type[top7_cols_location]

top7_cols_offender_race = df_offender_race.sum().sort_values(ascending=False).head(7).index
df_offender_race = df_offender_race[top7_cols_offender_race]

top7_offense_type = df_offense_type.sum().sort_values(ascending=False).head(7).index
df_offense_type = df_offense_type[top7_offense_type]

top7_victim_type = df_victim_type.sum().sort_values(ascending=False).head(7).index
df_victim_type = df_victim_type[top7_victim_type]

top7_bias = df_bias.sum().sort_values(ascending=False).head(7).index
df_bias = df_bias[top7_bias]

top7_bias_category = df_bias_category.sum().sort_values(ascending=False).head(7).index
df_bias_category = df_bias_category[top7_bias_category]
```

EXPLORATORY DATA ANALYSIS

Creating a Hate Crime Database

```
hateCrimeDatabase = pd.concat([
    df_location_type,
    df_offender_race,
    df_offense_type,
    df_victim_type,
    df_bias,
    df_bias_category
], axis=1)
hateCrimeDatabase.head(20)
hateCrimeDatabase['Monthly Count'] = df_monthly_totals
hateCrimeDatabase.head(20)
```

- Combined 6 key datasets into one master DataFrame
- Added monthly total counts as a reference column
- Prepared a clean, unified table for further statistical analysis

	Residence/Home	Highway/Road/Alley/Street/Sidewalk	School-Elementary/Secondary	Other/Unknown	Parking/Drop Lot/Garage	Church/Synagogue/Temple/Mosque	Park
01-2022	433	238	147	198	90	59	
02-2022	495	247	192	230	92	60	
03-2022	496	280	177	180	100	67	
04-2022	559	354	191	203	126	77	
05-2022	689	395	144	274	129	107	
06-2022	662	411	54	244	131	123	

Black or African American	Unknown	...	Anti-Hispanic or Latino	Anti-Lesbian, Gay, Bisexual, or Transgender (Mixed Group)	Anti-Asian	Race/Ethnicity/Ancestry	Religion	Sexual Orientation	Gender Identity	Disability	Gender	Monthly Count
310	320	...	94	65	76	1016	306	276	67	30	14	882
364	358	...	107	85	104	1161	370	303	74	27	14	1077
360	295	...	120	94	89	1136	367	316	94	30	17	1104
375	361	...	139	104	99	1262	421	355	113	27	23	1058
419	422	...	147	187	106	1341	464	510	112	25	30	1426

EXPLORATORY DATA ANALYSIS

Trend of Total Hate Crime Incidents Over Time

```
plt.figure(figsize=(10,5))
plt.plot(hateCrimeDatabase.index, hateCrimeDatabase['Monthly Count'],
         color='blue', label='Total Incidents')
plt.title('Total Incidents Over Time')
plt.xlabel('Month')
plt.ylabel('Count')
plt.legend()
plt.grid(True)
plt.xticks(
    ticks=range(0, len(hateCrimeDatabase.index), 12),
    labels=hateCrimeDatabase.index[::12],
    rotation=45
)
plt.show()
```

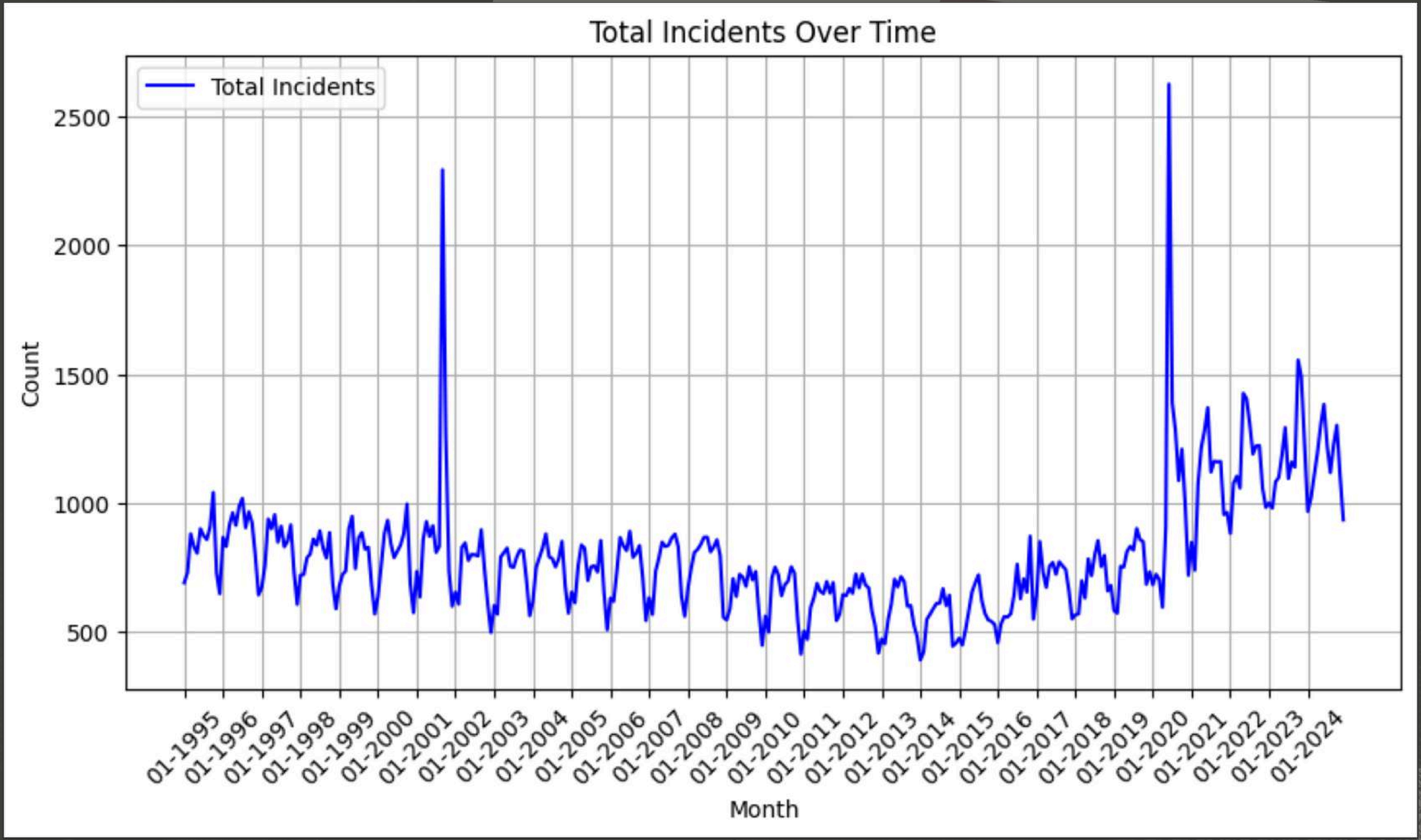
- Created a line chart showing the total number of hate crimes per month.
- The plot reveals how hate crimes fluctuate over time, showing visible ups and downs rather than a flat trend.
- Serves as an overview before testing the hypotheses.

DATA > OPINION

Question for the Audience

Why do you think there are spikes in the following graph on the next page?

EDA


Total Incidents Over Time
Total Incidents
Count
2500
2000
1500
1000
500
01-1995
01-1996
01-1997
01-1998
01-1999
01-2000
01-2001
01-2002
01-2003
01-2004
01-2005
01-2006
01-2007
01-2008
01-2009
01-2010
01-2011
01-2012
01-2013
01-2014
01-2015
01-2016
01-2017
01-2018
01-2019
01-2020
01-2021
01-2022
01-2023
01-2024
Month

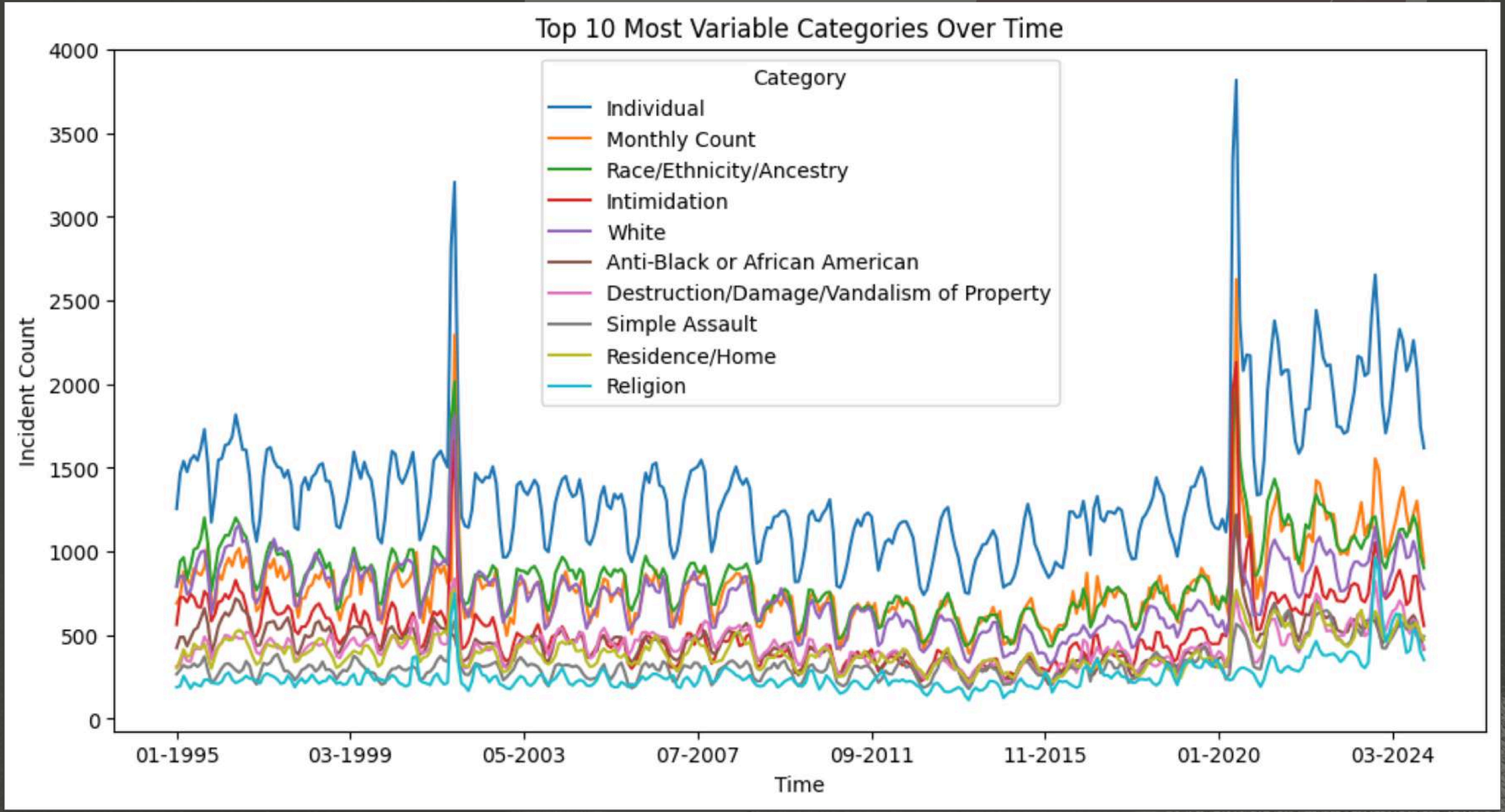
EDA

Top 10 Most Variable Hate Crime Categories Over Time

```
top_vars = hateCrimeDatabase.var().sort_values(ascending=False).head(10).index
hateCrimeDatabase[top_vars].plot(figsize=(12,6))
plt.title("Top 10 Most Variable Categories Over Time")
plt.xlabel("Time")
plt.ylabel("Incident Count")
plt.legend(title="Category")
plt.show()
```

- Calculated variance for every hate-crime related column.
- Selected the top 10 categories with the greatest month-to-month changes.
- Plotted them to observe which types of hate crimes show the most volatility.
- Reveals dynamic categories where hate incidents rise or fall sharply over time.

EXPLORATORY DATA ANALYSIS



RESEARCH QUESTION 1

- **Research question**: How does the frequency of hate crimes between racial or ethnic groups in the US differ over time.
- **Null Hypothesis (H_0)**: There is no significant difference in the frequency of hate crimes among different racial or ethnic groups over time.
- **Alternative Hypothesis (H_1)**: There is a significant difference in the frequency of hate crimes among different racial or ethnic groups over time.

ANOVA

```python
race_columns = [
    'White',
    'Black or African American',
    'Asian',
    'American Indian or Alaska Native',
    'Multiple',
    'Native Hawaiian or Other Pacific Islander'
]
df_long = hateCrimeDatabase[race_columns].melt(
    var_name='Race/Ethnicity',
    value_name='Incident Count')
groups = [df_long.loc[df_long['Race/Ethnicity'] == race, 'Incident Count'] for race in race_columns]
f_stat, p_value = stats.f_oneway(*groups)
print("ANOVA test results:")
print(f"F-statistic: {f_stat:.3f}")
print(f"P-value: {p_value:.6f}")
alpha = 0.05
if p_value < alpha:
    print("Reject H₀ → Significant difference in hate crime frequency across racial/ethnic groups.")
else:
    print("Fail to reject H₀ → No significant difference between groups.")
```

ANOVA test results:

F-statistic: 3665.100

P-value: 0.000000

Reject H_0 → Significant difference in hate crime frequency across racial/ethnic groups.

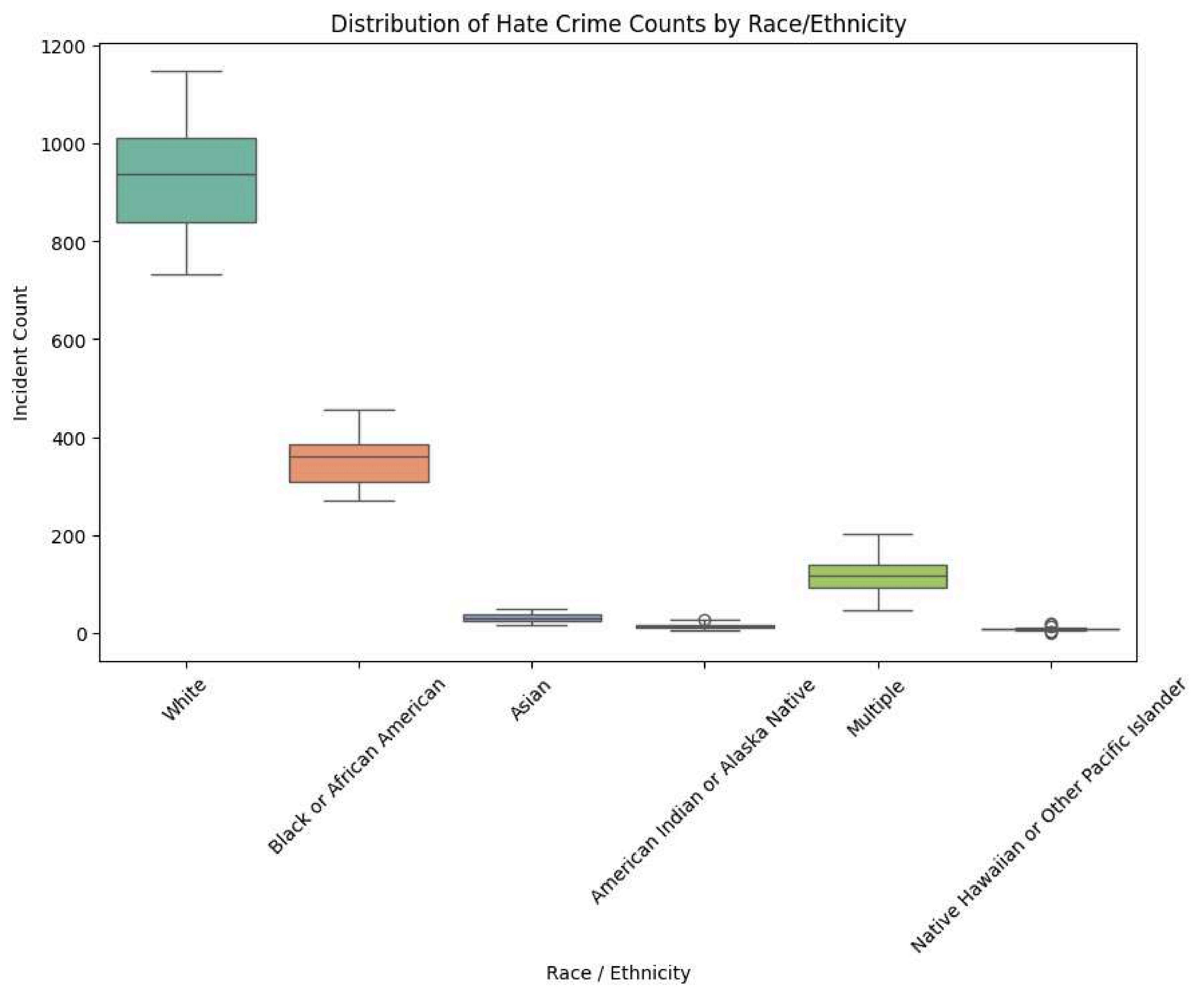
KRUSKAL WALLIS TEST

Similar to ANOVA but it doesnt assume a normal distribution, which is safer for our data set seeing as its a count based dataframe.

```
from scipy import stats

groups = [df_long.loc[df_long['Race/Ethnicity'] == race, 'Incident Count'] for race in race_columns]

h_stat, p_value = stats.kruskal(*groups)
print("Kruskal-Wallis H-test results:")
print(f"H-statistic: {h_stat:.3f}")
print(f"P-value: {p_value:.5f}")

if p_value < 0.05:
    print("Reject H₀ → Significant difference among groups.")
else:
    print("Fail to reject H₀ → No significant difference.")
```

```
Kruskal-Wallis H-test results:
H-statistic: 2005.482
P-value: 0.00000
Reject H₀ → Significant difference among groups.
```

VISUALIZATION WITH BOX PLOT



- White: highest median and widest spread, consistently high counts with large variability.
- Black or African American: second-highest median; moderate spread.
- Multiple: mid-level counts but quite variable (long whiskers).
- Asian and American Indian or Alaska Native: low medians, relatively tight spread.
- Native Hawaiian or Other Pacific Islander: lowest counts overall.

RESEARCH QUESTION 2

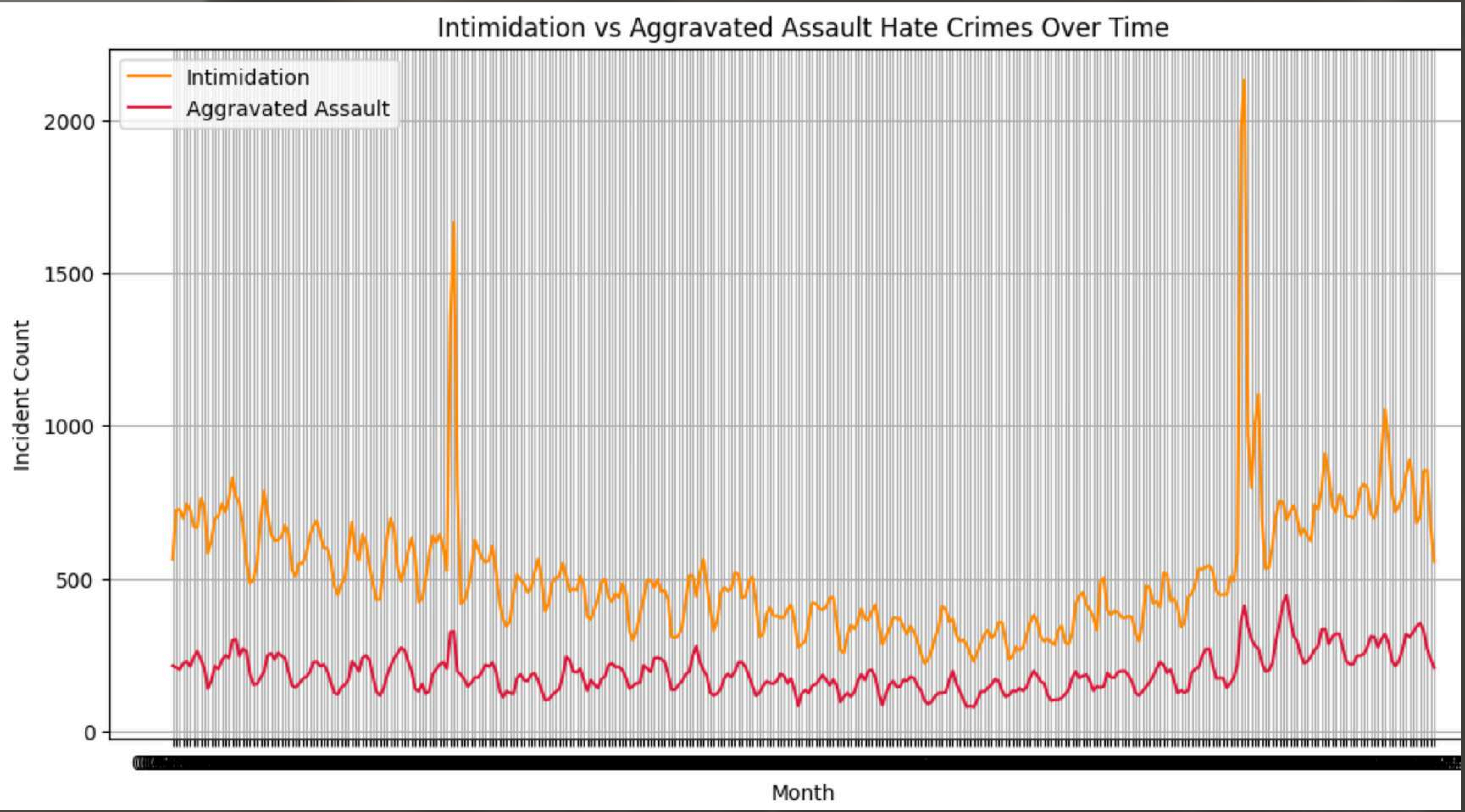
Research Question: How does the relationship between *intimidation* and *aggravated assault change* over time?

- **Null Hypothesis (H_0)**: There is no correlation between intimidation and aggrevated assault over time.
- **Alternative Hypothesis (H_1)**: Increases in intimidation hate crimes also leads to increases of aggravated assault over time.

Question for the Audience

Are words just words, or do you think words are a precursor for action?

INTIMIDATION VS AGGRAVATED ASSAULTS: MONTHLY TRENDS



- Both series rise and fall in broadly similar periods.
- Intimidation is much larger in magnitude;
  aggravated assault tracks the ups/downs with a smaller amplitude.

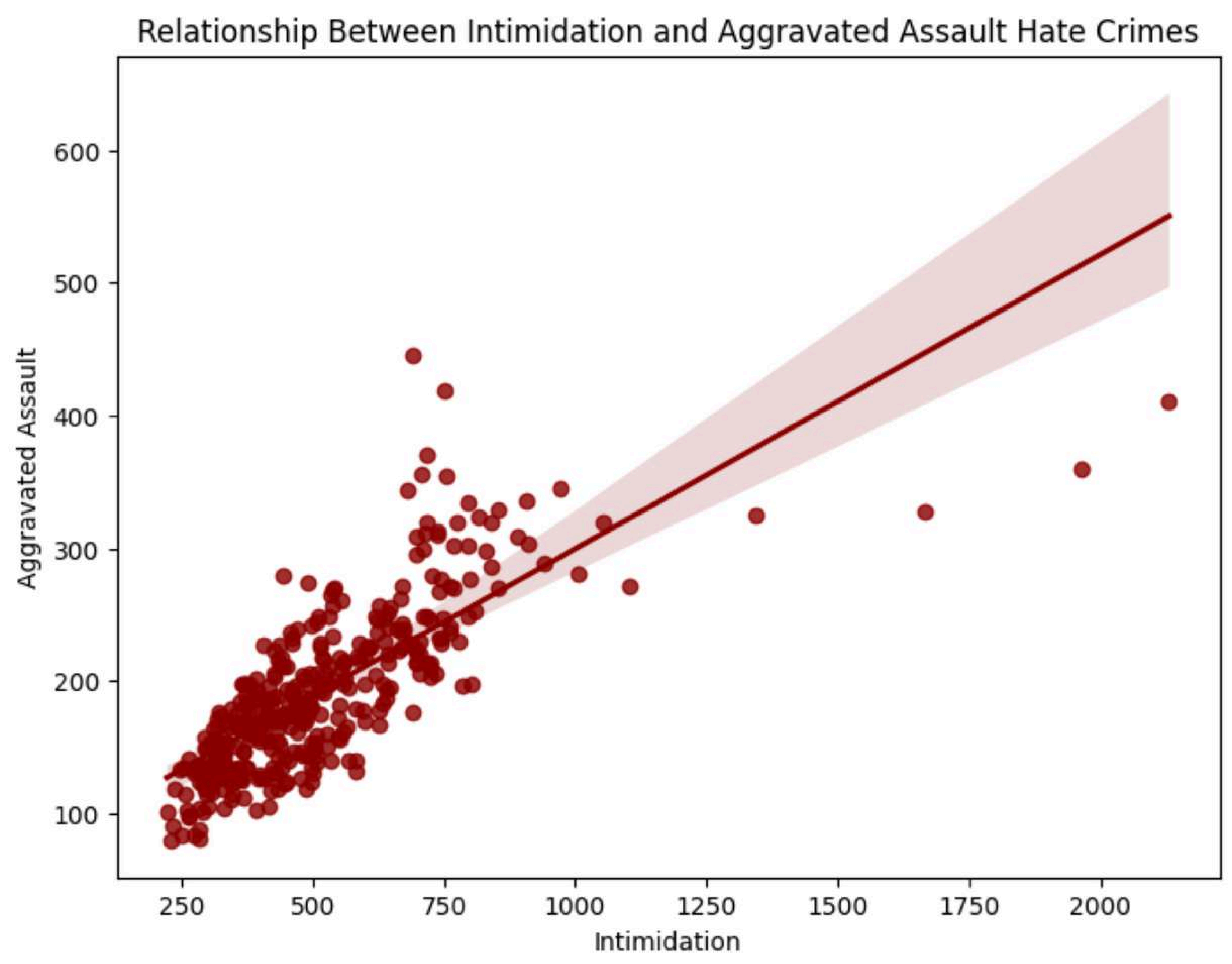
CORRELATION BETWEEN INTIMIDATION AND AGGRAVATED ASSAULTS

```
corr = df_relation['Intimidation'].corr(df_relation['Aggravated Assault'])
print("Pearson correlation:", corr)

from scipy.stats import pearsonr
r, p = pearsonr(df_relation['Intimidation'], df_relation['Aggravated Assault'])
print(f"r = {r:.3f}, p-value = {p:.4f}")


import seaborn as sns
plt.figure(figsize=(8,6))
sns.regplot(x='Intimidation', y='Aggravated Assault', data=df_relation, color='darkred')
plt.title('Relationship Between Intimidation and Aggravated Assault Hate Crimes')
plt.show()
```

r = 0.764, p-value = 0.0000

Reject Null Hypothesis (H_0): There is no correlation between intimidation and aggravated assault over time.

CORRELATION BETWEEN INTIMIDATION AND AGGRAVATED ASSAULTS



- Upward trend: The scatter shows a clear positive slope.
- Pearson's r ≈ 0.76 → strong positive association.
- Significance: p ≈ 0.000 (< 0.05) → reject H_0
- Months with higher Intimidation tend to coincide with higher Aggravated Assault counts.

RESEARCH QUESTION 3

Research Question: Has the overall number of reported hate crimes fluctuated significantly, indicating temporal or seasonal trends?

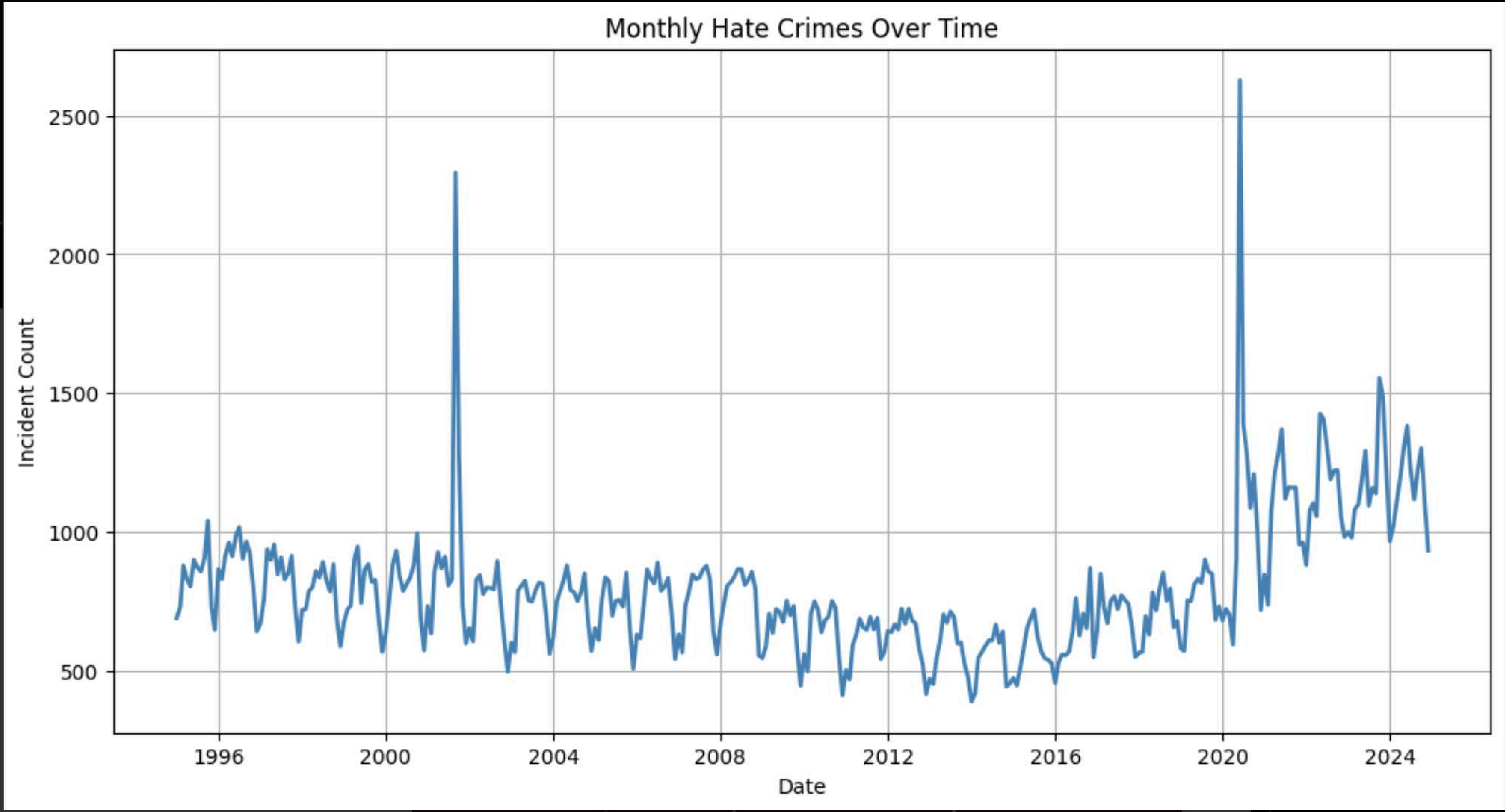
- **Null Hypothesis (H_0)**: The number of hate crimes has remained stable over time, showing no significant temporal or seasonal variation.

- **Alternative Hypothesis (H_1)**: The number of hate crimes has varied significantly over time, showing clear temporal or seasonal fluctuations.

Question for the Audience

Do you think the number of hate crimes in the US in the past 30 years has increased or decreased?

GROUND



HE INDUSTRY'S HISTORY

3.1 Linear Trend Analysis (OLS): Establishes overall upward trend in hate crimes.

```python
import statsmodels.api as sm
import numpy as np


hateCrimeDatabase = hateCrimeDatabase.sort_index()
hateCrimeDatabase['Time_Index'] = np.arange(len(hateCrimeDatabase))


X = sm.add_constant(hateCrimeDatabase['Time_Index'])
y = hateCrimeDatabase['Monthly_Count']
model = sm.OLS(y, X).fit()
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          Monthly_Count   R-squared:                       0.046
Model:                            OLS   Adj. R-squared:                  0.044
Method:                 Least Squares   F-statistic:                     17.42
Date:                Sat, 25 Oct 2025   Prob (F-statistic):           3.76e-05
Time:                        07:58:05   Log-Likelihood:                -2476.9
No. Observations:                 360   AIC:                             4958.
Df Residuals:                     358   BIC:                             4966.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const        704.9609     24.832     28.389      0.000     656.125     753.797
Time_Index     0.4997      0.120      4.174      0.000       0.264       0.735
==============================================================================
Omnibus:                      211.854   Durbin-Watson:                   0.598
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2714.600
Skew:                           2.202   Prob(JB):                         0.00
Kurtosis:                      15.711   Cond. No.                         414.
==============================================================================
```

R square = 0.46

F-statistic = 17.42

p = 3.76e-05

There is a small but statistically relevant increase in hate crimes over the past 30 years

3.2 Temporal & Seasonal ANOVA: Tests yearly and monthly differences, confirming significant fluctuations.

```python
hateCrimeDatabase = hateCrimeDatabase.rename(columns={'Monthly Count': 'Monthly_Count'})
anova = sm.stats.anova_lm(
    sm.OLS.from_formula('Monthly_Count ~ C(Year)', data=hateCrimeDatabase).fit(),
    typ=2)
print(anova)
```

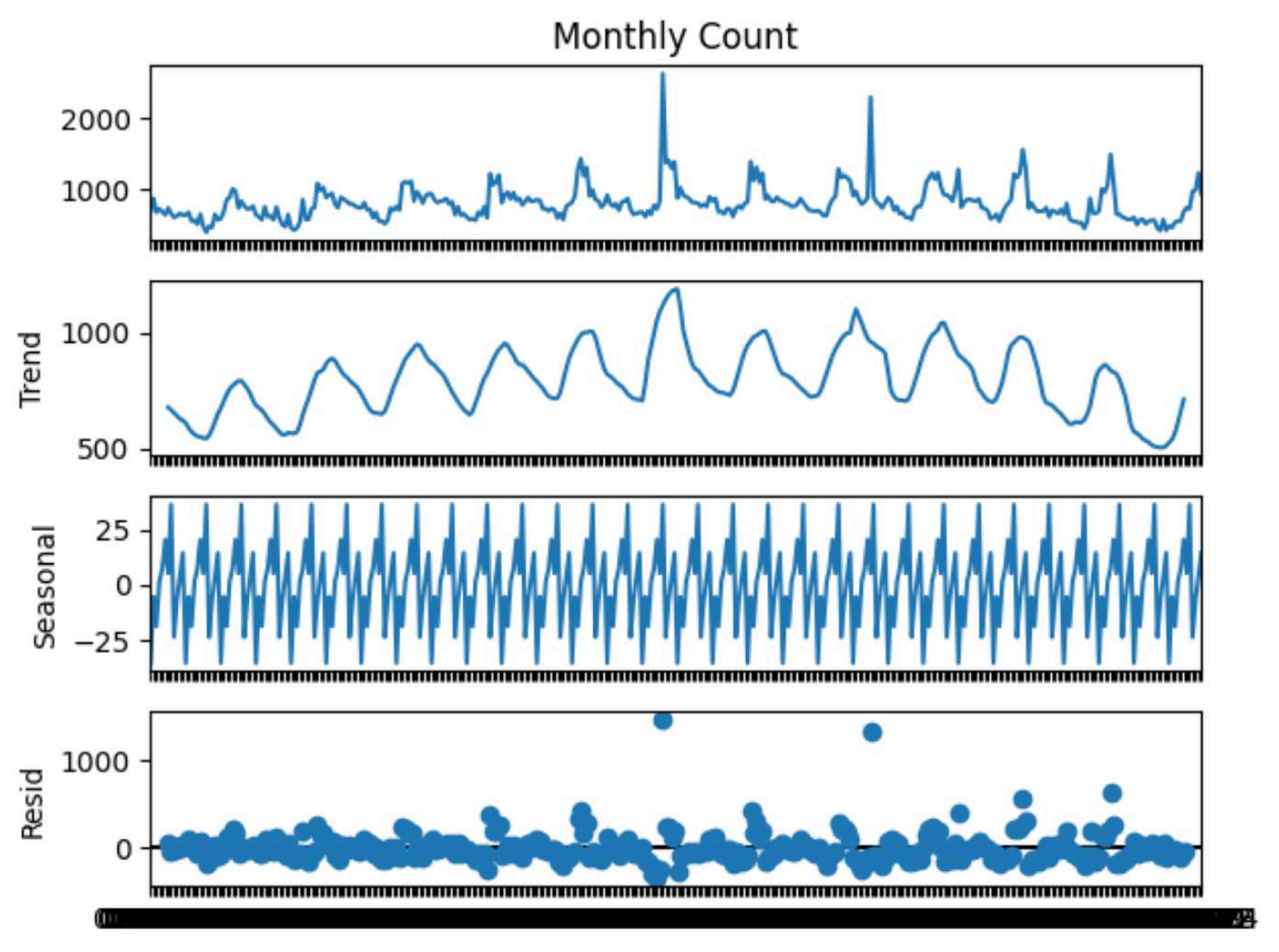
```
                sum_sq    df         F    PR(>F)
C(Year)    8510.888889   2.0  0.162699  0.850524
Residual 863124.083333  33.0       NaN       NaN
```

```python
hateCrimeDatabase['Month'] = hateCrimeDatabase.index.month
anova_month = sm.stats.anova_lm(
    sm.OLS.from_formula('Monthly_Count ~ C(Month)', data=hateCrimeDatabase).fit(),
    typ=2)
print(anova_month)
```

```
                sum_sq    df         F    PR(>F)
C(Month) 562211.638889  11.0  3.964289  0.002319
Residual 309423.333333  24.0       NaN       NaN
```

ROUND

HE INDUSTRY'S HISTORY



There's a long-term upward trend (confirmed by OLS).

There are significant seasonal patterns (confirmed by ANOVA on months).

There are irregular spikes likely tied to major real-world events (reflected in residuals).

HE INDUSTRY'S HISTORY

3.3 Multivariate Time-Series Modeling (VAR): Examines dynamic interrelationships between hate crime categories.

```
cols_offense = [
    'Intimidation',
    'Destruction/Damage/Vandalism of Property',
    'Simple Assault',
    'Aggravated Assault',
    'Robbery',
    'Burglary/Breaking & Entering',
    'All Other Larceny',
    'Monthly_Count'
]

from statsmodels.tsa.api import VAR

# Example: Offense pattern VAR
df = hateCrimeDatabase[cols_offense].dropna()

# Stationarity check & differencing
df_diff = df.diff().dropna()

# Fit
model = VAR(df_diff)
results = model.fit(ic='aic')
print(results.summary())
```

This model analyzes how multiple variables move together over time
The response is way too large to add to this presentation, but we added the final matrix correlation for useful insights.

Variable Pair	Correlation	Interpretation
Intimidation ⇔ Destruction/Vandalism	0.6	Moderate–strong — these still move together even after the model, suggesting similar causes or missing shared drivers (like social tensions).
Intimidation ⇔ Simple Assault	0.4	Moderate correlation — possibly escalation patterns (intimidation often co-occurs with assault).
Intimidation ⇔ Aggravated Assault	0.47	Similar trend — both forms of aggression remain linked.
Destruction/Vandalism ⇔ Simple Assault	0.52	Strong — both may reflect general unrest or hate spikes.
Destruction/Vandalism ⇔ Aggravated Assault	0.39	Still somewhat connected.

Correlation matrix of residuals

	Intimidation	Destruction/Damage/Vandalism of Property	Simple Assault	Aggravated Assault	Robbery	Burglary/Breaking & Entering	All Other Larceny	Monthly_Count
Intimidation	1.000000	0.599305	0.396815	0.467995	0.167291	0.115167	0.136115	0.079372
Destruction/Damage/Vandalism of Property	0.599305	1.000000	0.515267	0.388177	0.225308	0.130226	0.153577	0.059330
Simple Assault	0.396815	0.515267	1.000000	0.408311	0.143351	0.164919	-0.003512	0.099946
Aggravated Assault	0.467995	0.388177	0.408311	1.000000	0.132382	0.197237	0.107475	0.069482
Robbery	0.167291	0.225308	0.143351	0.132382	1.000000	-0.029988	-0.045358	-0.003183
Burglary/Breaking & Entering	0.115167	0.130226	0.164919	0.197237	-0.029988	1.000000	0.087447	-0.012090
All Other Larceny	0.136115	0.153577	-0.003512	0.107475	-0.045358	0.087447	1.000000	0.201246
Monthly_Count	0.079372	0.059330	0.099946	0.069482	-0.003183	-0.012090	0.201246	1.000000

POSSIBLE IMPROVEMENTS AND IMPLICATIONS

1) Add current affairs and social data to try to explain gaps not explained by models.
2) Include more data points over more years to improve accuracy
3) Remove seasonality to understand relationships between variables more clearly

1) Hate crimes are a persistent and growing social issue, and policymakers should continue to push to solve this issue
2) Future studies should incorporate external variables.

CONCLUSION

RSQ1: Reject H_0 which says that there is a significant difference in hate crime frequency across racial/ethnic groups.

RSQ2: Reject Null Hypothesis (H_0), so it means that there is correlation between intimidation and aggravated assault over time.

RSQ3: Reject Null Hypothesis (H_0) which means the number of hate crimes has varied stable over time, showing no significant temporal or seasonal variation.

REFERENCES

1. Federal Bureau of Investigation. (n.d.). Crime Data Explorer (CDE). U.S. Department of Justice. Retrieved October 25, 2025, from https://cde.ucr.cjis.gov/
2. Federal Bureau of Investigation. (2023). Uniform Crime Reporting (UCR) Program. Retrieved from https://www.fbi.gov/how-we-can-help-you/more-fbi-services-and-information/ucr
3. Federal Bureau of Investigation. (2023, February 8). Privacy Impact Assessment for the Uniform Crime Reporting (UCR) System. U.S. Department of Justice. Retrieved from https://www.fbi.gov/file-repository/pias/pia-uniform-crime-reporting-system-020823.pdf
4. Federal Bureau of Investigation. (n.d.). Hate Crime Statistics Program Methodology. Retrieved from https://ucr.fbi.gov/hate-crime/2018/resource-pages/methodology

WE WANT TO SAY

THANK YOU

FOR YOUR ATTENTION