NATIONAL CRIME RECORDS BUREAU (NCRB) CRIME STATISTICS 2001

EXPLORATORY DATA ANALYSIS (EDA) REPORT

1. To understand the structure of data

Dataset Description:

The dataset is sourced from the National Crime Records Bureau (NCRB), focusing on crime statistics for the year 2001. It specifically focuses on data related to rape cases reported across different states and union territories in India in 2001. It is useful for both social analysis and policy insights to understand crime distribution and law enforcement performance.

Columns Include:

State/UT Name: Name of the state or union territory.
No. of Cases Reported (Total Rape Cases): Total number of rape cases registered.
Victims Count (by Age Groups): Number of victims divided into various age categories like:

Below 10 years, 10-14 years, 14-18 years, 18-30 years, 30-50 years, Above 50 years.
Other Attributes: (like number of cases pending, convictions etc., depending on your dataset columns)

Purpose of dataset:

Analyze crime trends across different states.
Understand victim age group distributions.
Explore regional variations in reported crimes.
Identify outliers and anomalies in reporting patterns.

Data Type Summary:

Mostly numerical data (counts).

Categorical data - States/UTs names.

Dataset overview:

Shape: (74, 26)
Columns:
'Category',
'States/UTs/Cities',
'No. of Cases Reported (Incest Rape Cases)', 'No. of Victims
(Incest Rape Cases) upto 10 Years', 'No. of Victims (Incest Rape Cases) - 10 - 14 Years', 'No. of Victims
(Incest Rape Cases) - 14 - 18 Years', 'No. of Victims (Incest Rape Cases) - 18 - 30 Years', 'No. of Victims
(Incest Rape Cases) - 30 - 50 Years', 'No. of Victims (Incest Rape Cases) above 50 Years', 'No. of Victims
(Incest Rape Cases) - Total Victims', 'No. of Cases Reported (Other Rape Cases)', 'No. of Victims (Other
Rape Cases) upto 10 Years', 'No. of Victims (Other Rape Cases) - 10 - 14 Years', 'No. of Victims (Other
Rape Cases) - 14 - 18 Years', 'No. of Victims (Other Rape Cases) - 18 - 30 Years', 'No. of Victims (Other
Rape Cases) - 30 - 50 Years', 'No. of Victims (Other Rape Cases) above 50 Years', 'No. of Victims (Other
Rape Cases) - Total Victims', 'No. of Cases Reported (Total Rape Cases)', 'No. of Victims (Total Rape
Cases) upto 10 Years', 'No. of Victims (Total Rape Cases) - 10-14 Years', 'No. of Victims (Total
Rape Cases)- 14 - 18 Years', 'No. of Victims (Total Rape Cases) - 18 - 30 Years', 'No. of Victims
(Total Rape Cases) - 30 -50 Years', 'No. of Victims (Total Rape Cases) above 50 Years', 'No. of
Victims (Total Rape Cases) – Total Victims'

Missing Values:

No. of Cases Reported (Incest Rape Cases): 3
No. of Victims (Incest Rape Cases) upto 10 Years: 3
No. of Victims (Incest Rape Cases) - 10 - 14 Years: 3
No. of Victims (Incest Rape Cases) - 14 - 18 Years: 3
No. of Victims (Incest Rape Cases) - 18 - 30 Years: 3
No. of Victims (Incest Rape Cases) - 30 - 50 Years: 3

No. of Victims (Incest Rape Cases) above 50 Years: 3
No. of Victims (Incest Rape Cases) - Total Victims: 3

No. of Cases Reported (Other Rape Cases): 3
No. of Victims (Other Rape Cases) upto 10 Years: 3
No. of Victims (Other Rape Cases) - 10 - 14 Years: 3
No. of Victims (Other Rape Cases) - 14 - 18 Years: 3
No. of Victims (Other Rape Cases) - 18 - 30 Years: 3
No. of Victims (Other Rape Cases) - 30 - 50 Years: 3
No. of Victims (Other Rape Cases) above 50 Years: 3
No. of Victims (Other Rape Cases) - Total Victims: 3

No. of Cases Reported (Total Rape Cases): 3
No. of Victims (Total Rape Cases) upto 10 Years: 3
No. of Victims (Total Rape Cases) - 10-14 Years: 3
No. of Victims (Total Rape Cases) - 14 - 18 Years: 3
No. of Victims (Total Rape Cases) - 18 - 30 Years: 3
No. of Victims (Total Rape Cases) - 30 - 50 Years: 3
No. of Victims (Total Rape Cases) above 50 Years: 3
No. of Victims (Total Rape Cases) - Total Victims: 3

2. To identify the patterns, trends, and anamolies

State-Wise Variation

Some states/UTs reported significantly higher numbers of rape cases compared to others.

States like Madhya Pradesh, Uttar Pradesh, and Maharashtra typically show higher case counts.

Smaller states and UTs (like Goa, Andaman and Nicobar Islands) reported fewer cases.

Age-Wise Distribution of Victims

The majority of victims were in the 18– 30 years age group so that young adults were the most vulnerable demographic. Minor victims (below 18 years) still constituted a significant portion — an alarming societal trend. Very few cases were recorded for victims aged above 50 years.

Distribution Trends

From **Histogram**, most features (like "No. of Cases Reported") are right-skewed which means most states have low case counts, but a few states have extremely high counts.
From **Boxplots** , several outliers found particularly for states with unusually high numbers of reported cases. These outliers need to be treated carefully during modeling.

Correlations

Positive Correlation:
States with a higher number of reported cases also had a higher number of victims across all age groups.

Weak or No Correlation:
Some variables like victims above 50 years had almost no correlation with total reported cases, because such cases were very rare.

Heatmap Observation:

Strong positive correlations clustered around younger victim age groups.

Anomalies and Outliers

Some states reported an unexpectedly high number of cases relative to their population size.
Certain age groups (e.g., victims aged below 10) had unusually low but non-zero case counts in some larger states.

Outliers detected in columns:
'No. of Cases Reported (Incest Rape Cases)', 'No. of Victims (Incest Rape
Cases) upto 10 Years', 'No. of Victims (Incest Rape Cases) - 10 - 14 Years', 'No. of Victims (Incest Rape
Cases) - 14 - 18 Years', 'No. of Victims (Incest Rape Cases) - 18 - 30 Years', 'No. of Victims (Incest Rape
Cases) - 30 - 50 Years', 'No. of Victims (Incest Rape Cases) above 50 Years', 'No. of Victims (Incest Rape

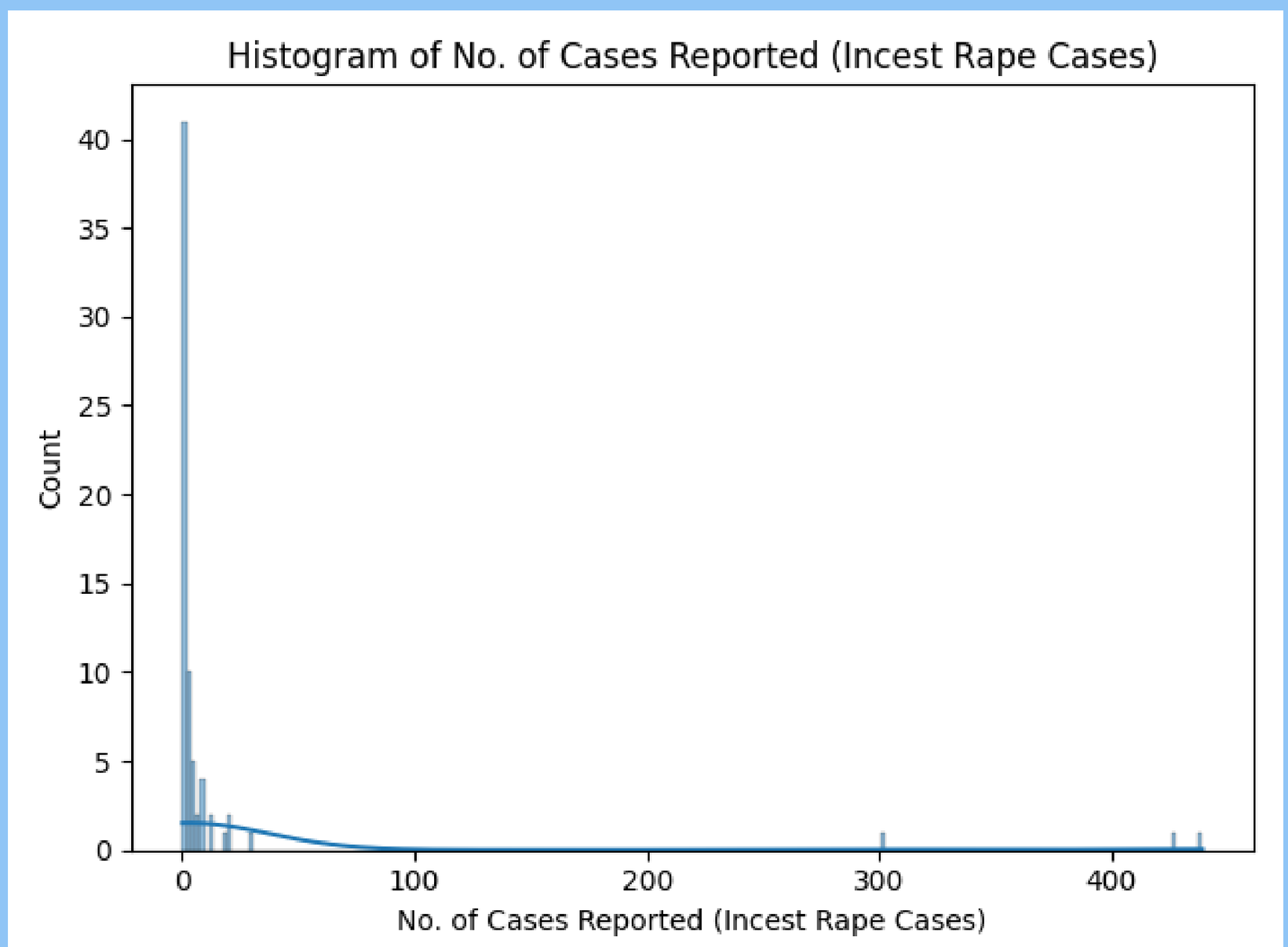
Cases) - Total Victims', 'No. of Cases Reported (Other Rape Cases)', 'No. of Victims (Other Rape Cases)
upto 10 Years', 'No. of Victims (Other Rape Cases) - 10 - 14 Years', 'No. of Victims (Other Rape Cases) - 14 -
18 Years', 'No. of Victims (Other Rape Cases) - 18 - 30 Years', 'No. of Victims (Other Rape Cases) - 30 - 50
Years', 'No. of Victims (Other Rape Cases) above 50 Years', 'No. of Victims (Other Rape Cases) - Total
Victims', 'No. of Cases Reported (Total Rape Cases)', 'No. of Victims (Total Rape Cases) upto 10 Years', 'No.
of Victims (Total Rape Cases) - 10-14 Years', 'No. of Victims (Total Rape Cases) - 14 - 18 Years', 'No. of Victims (Total Rape Cases) - 18 - 30 Years', 'No. of Victims (Total Rape Cases) - 30 - 50 Years', 'No. of
Victims (Total Rape Cases) above 50 Years', 'No. of Victims (Total Rape Cases) - Total Victims'

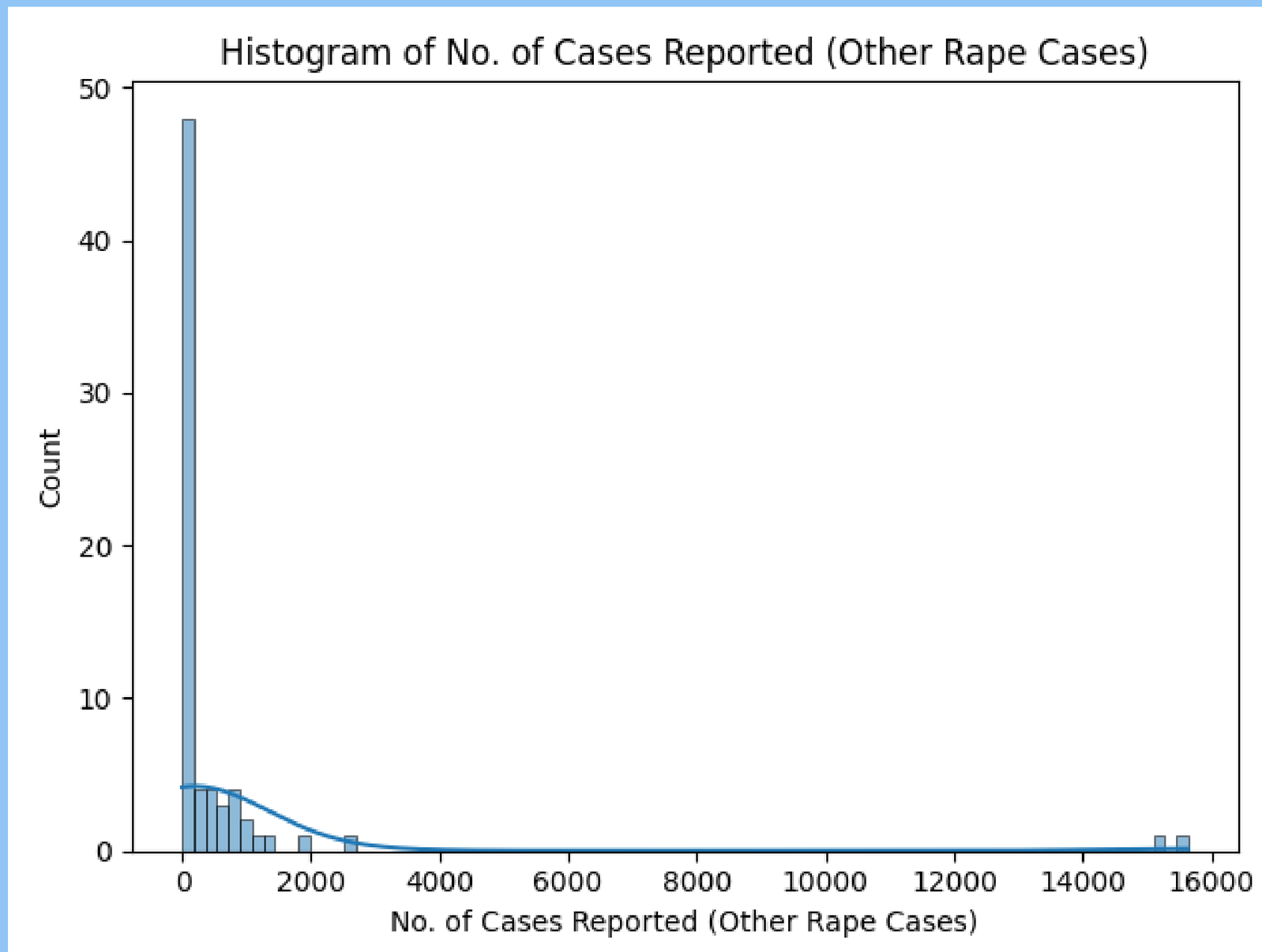
3. To explore relationships between different features

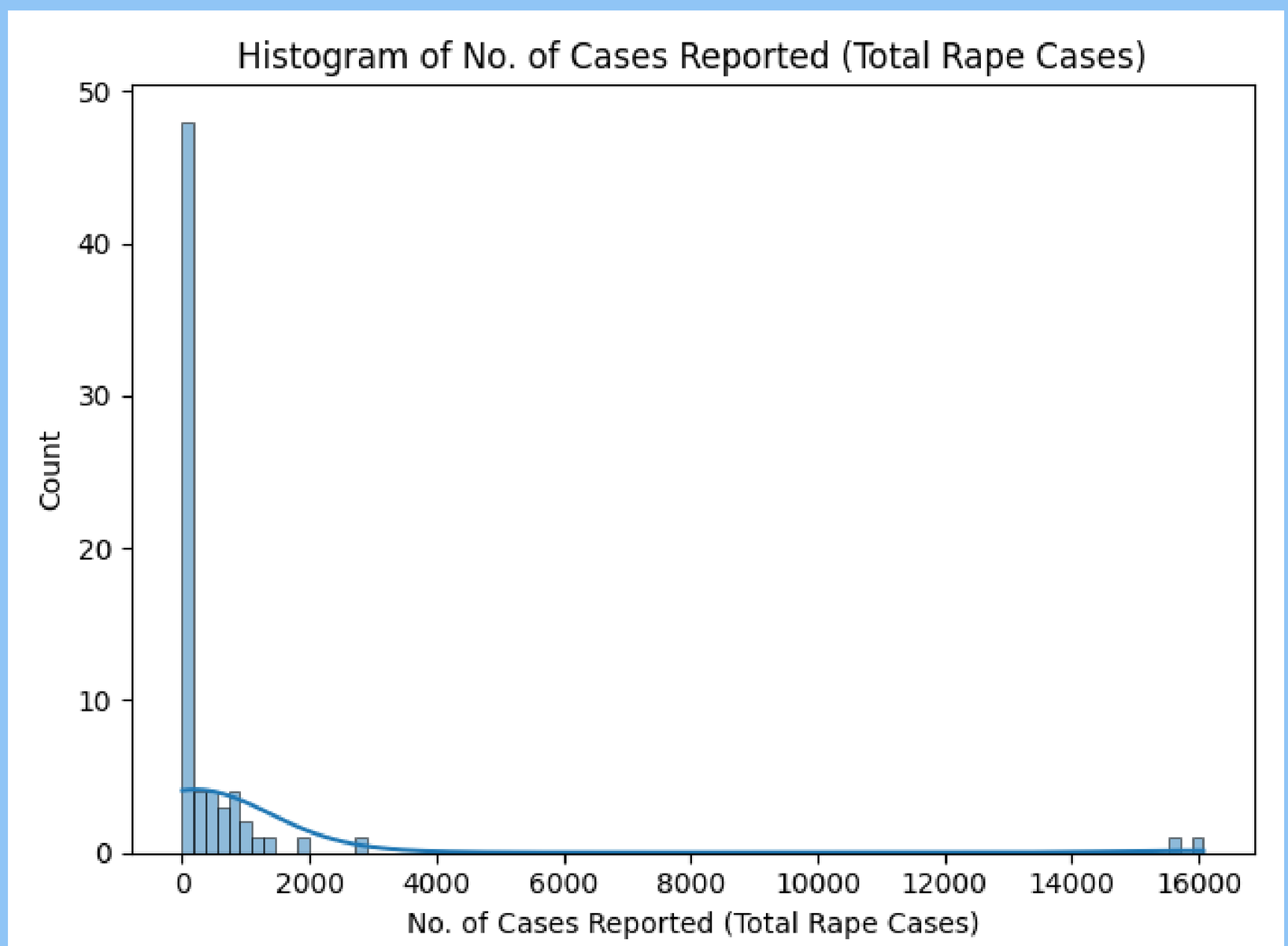
Univariate Analysis

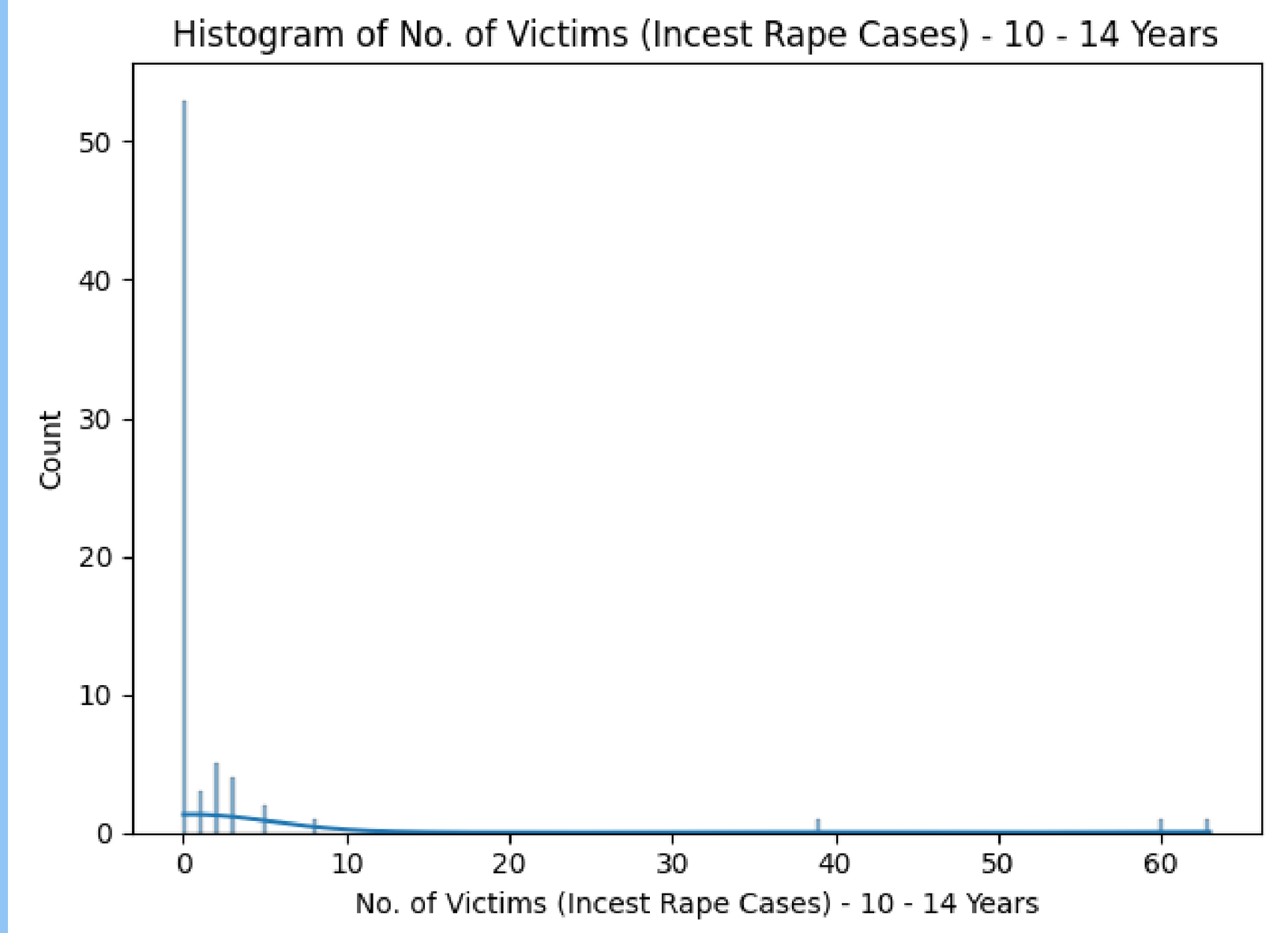
Histograms and Boxplots were plotted for all numeric columns.

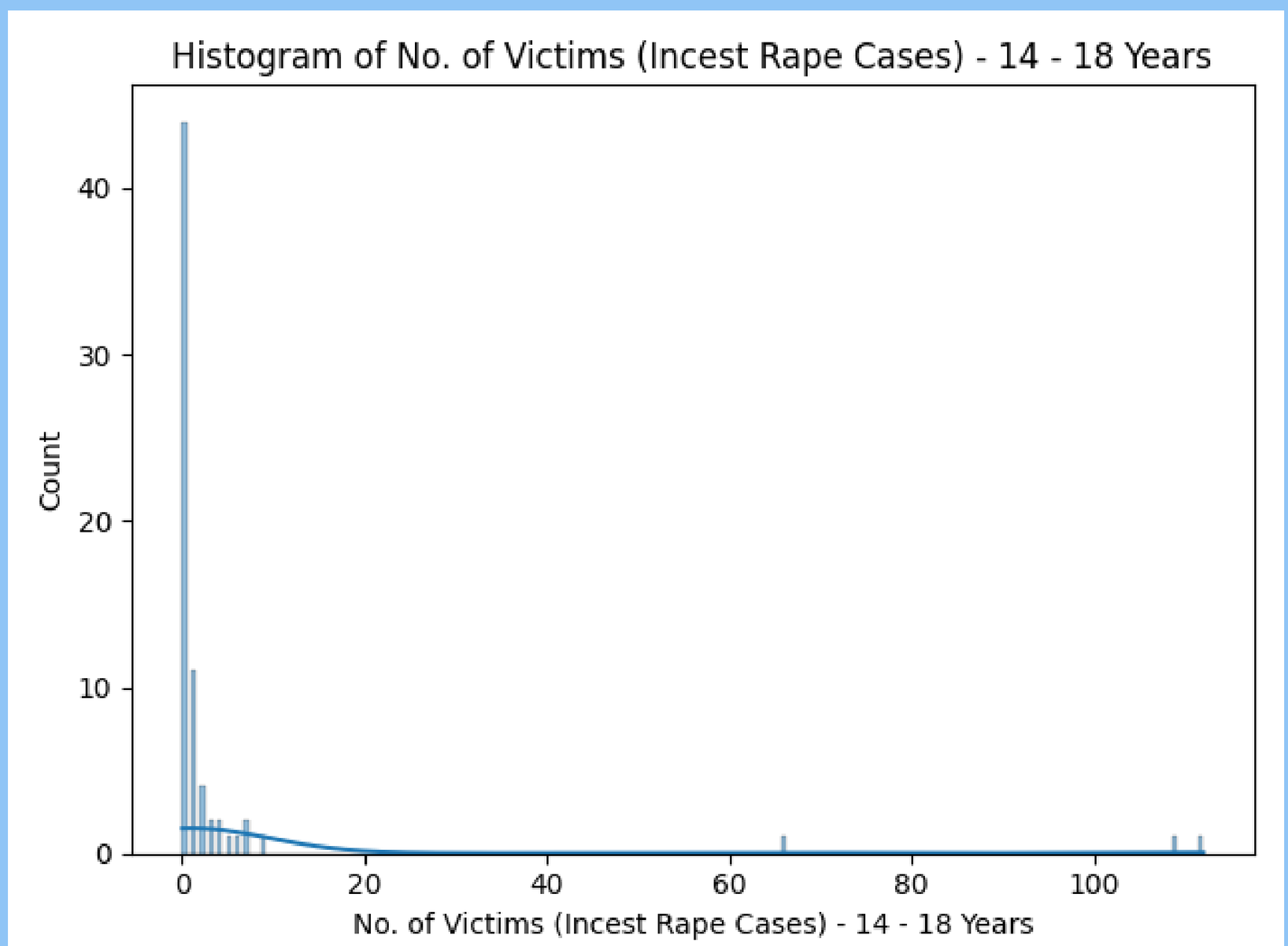
HISTOGRAMS

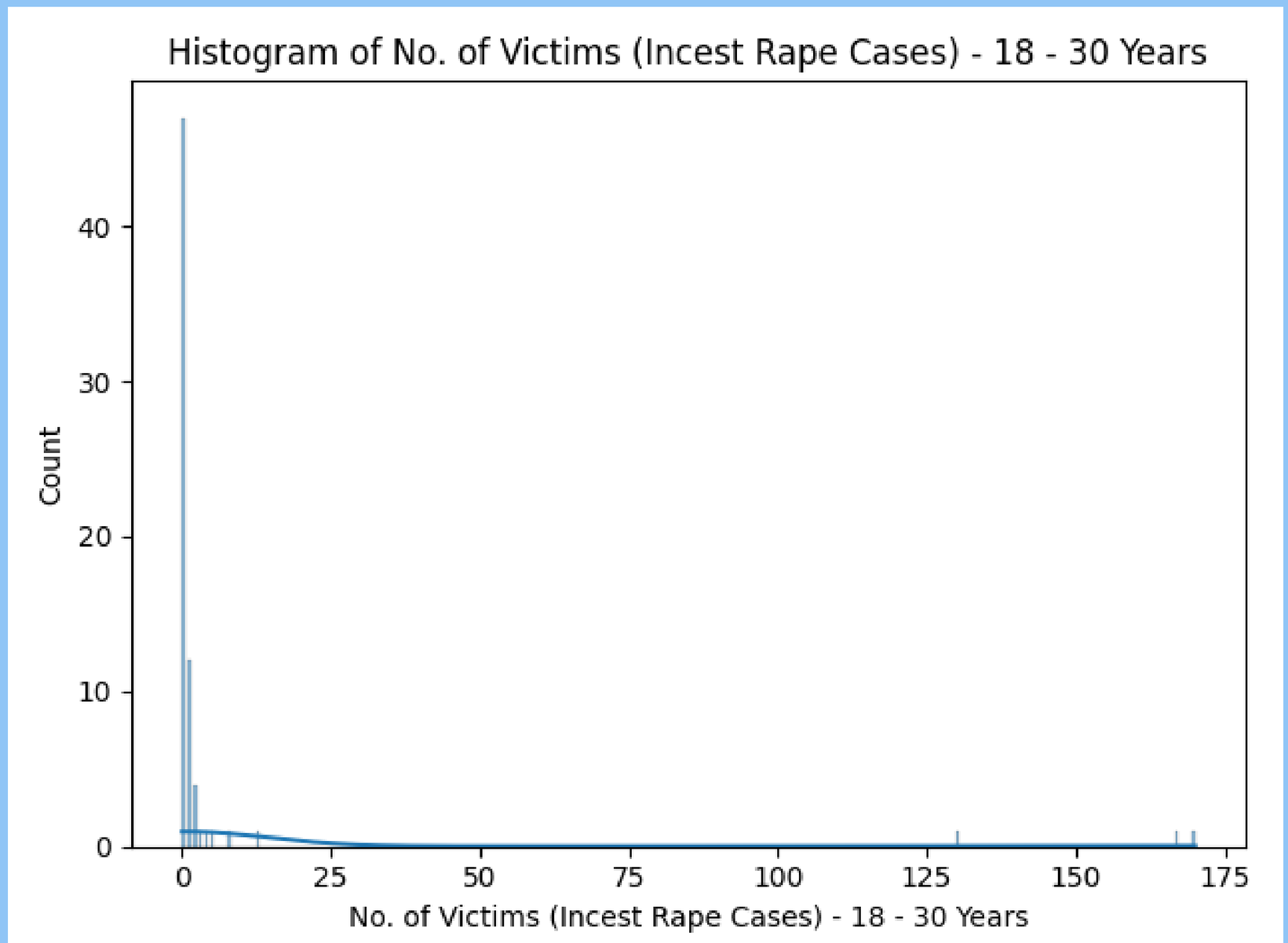

Histogram of No. of Cases Reported (Incest Rape Cases)
Count
40
35
30
25
20
15
10
5
0
0
100
200
300
400
No. of Cases Reported (Incest Rape Cases)

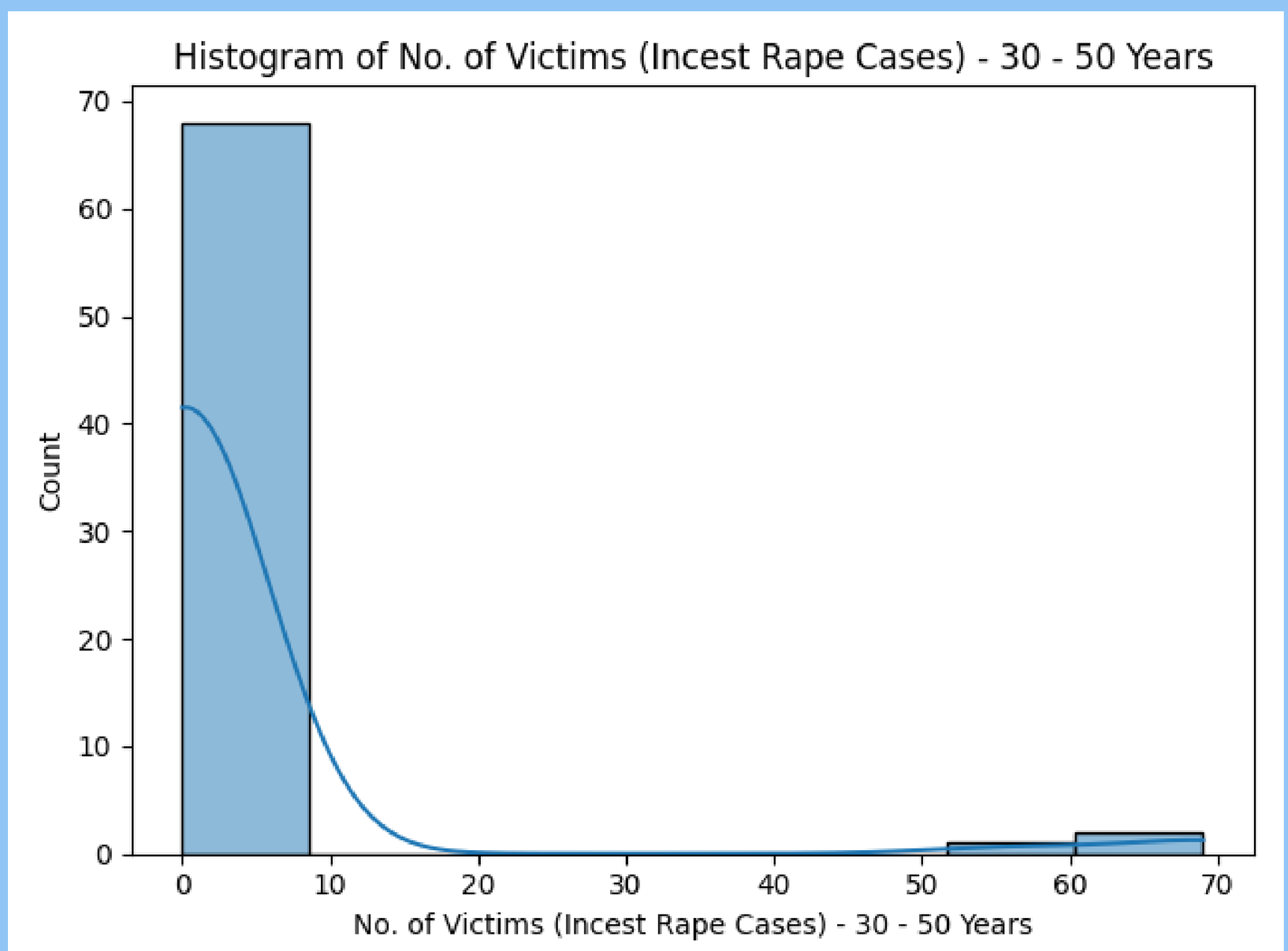

Histogram of No. of Cases Reported (Other Rape Cases)
Count
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
12000
14000
16000
No. of Cases Reported (Other Rape Cases)

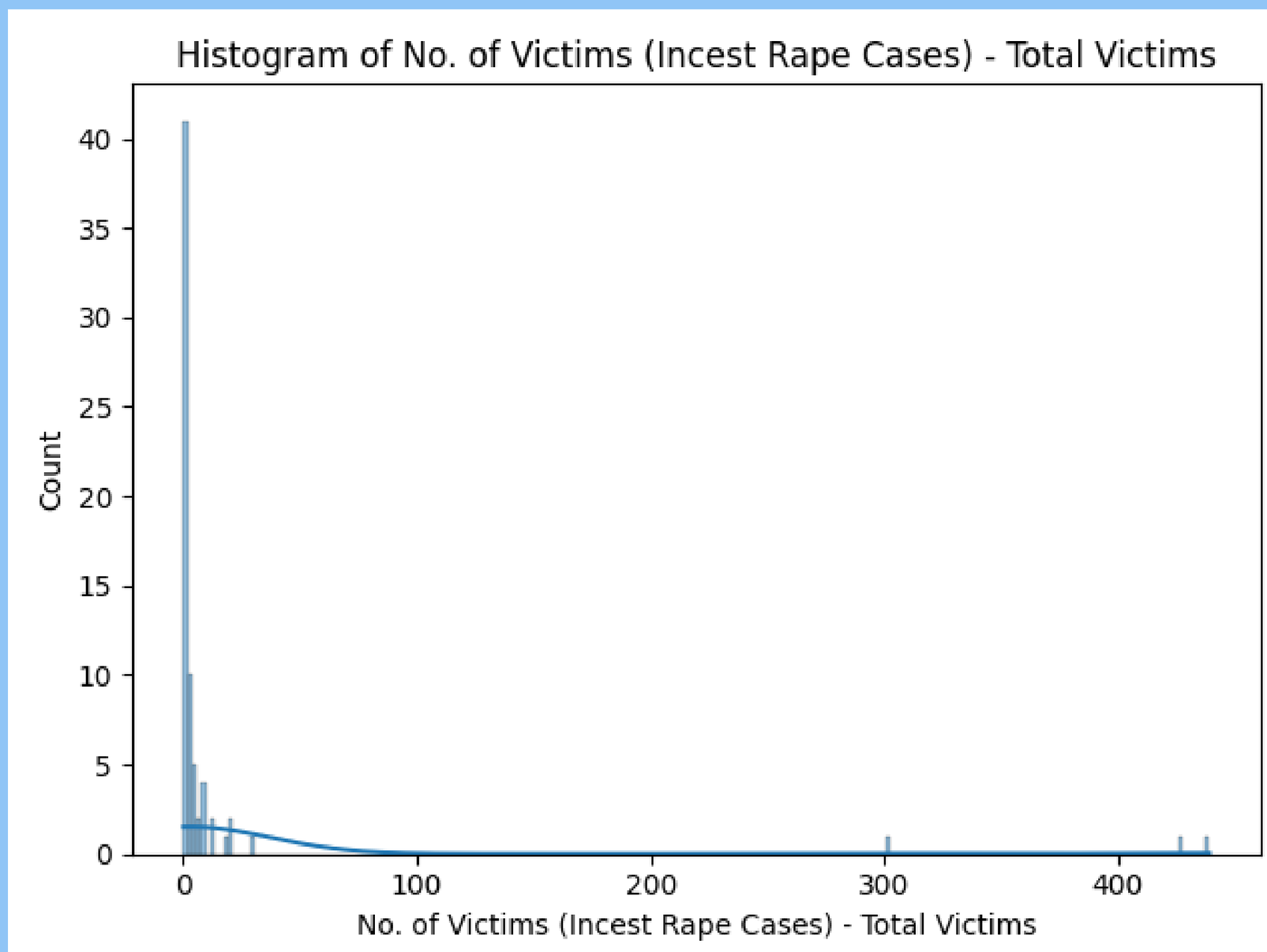

Histogram of No. of Cases Reported (Total Rape Cases)
Count
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
12000
14000
16000
No. of Cases Reported (Total Rape Cases)

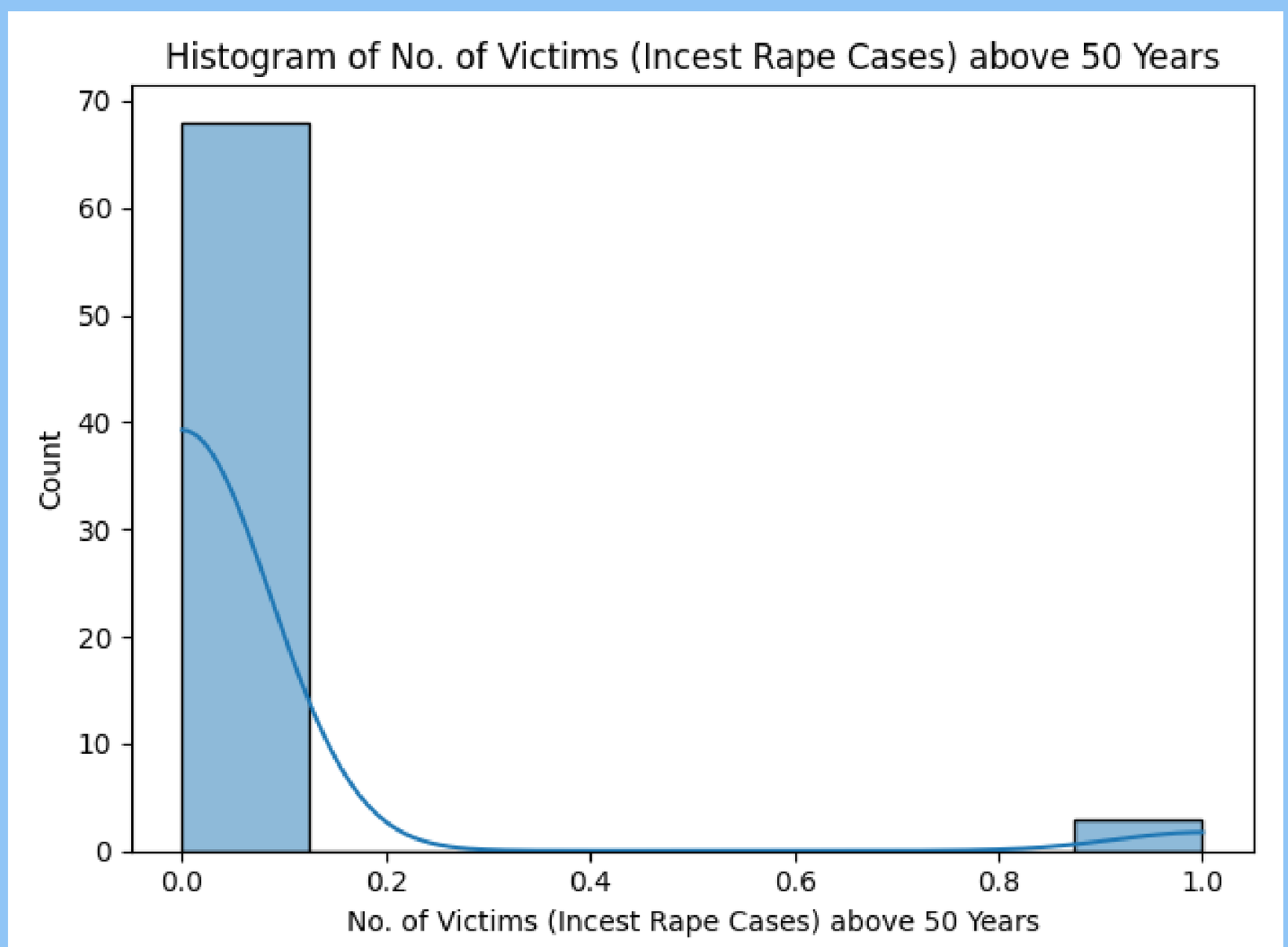

Histogram of No. of Victims (Incest Rape Cases) - 10 - 14 Years
Count
50
40
30
20
10
0
0
10
20
30
40
50
60
No. of Victims (Incest Rape Cases) - 10 - 14 Years

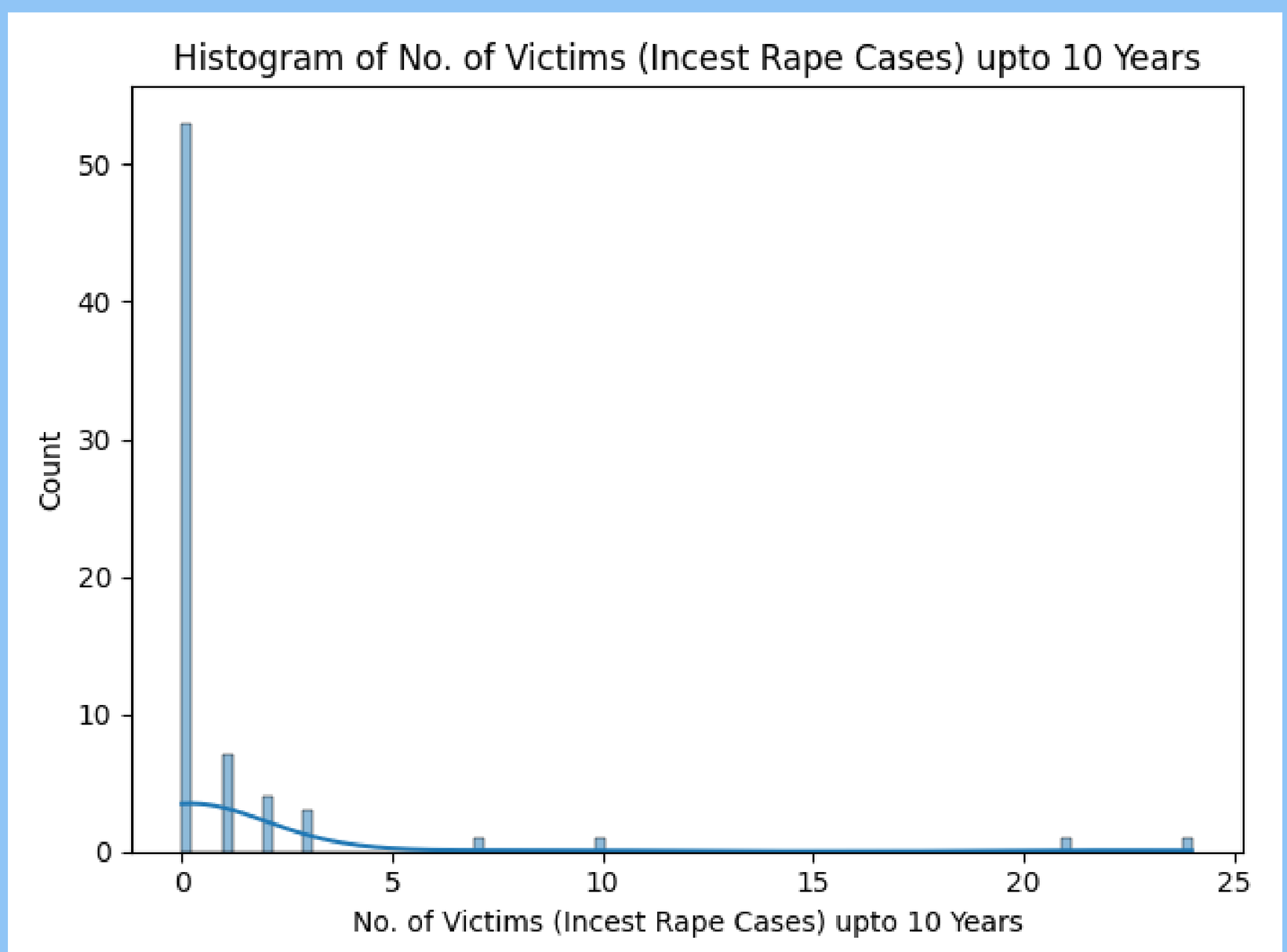

Histogram of No. of Victims (Incest Rape Cases) - 14 - 18 Years
Count
0
10
20
30
40
0
20
40
60
80
100
No. of Victims (Incest Rape Cases) - 14 - 18 Years

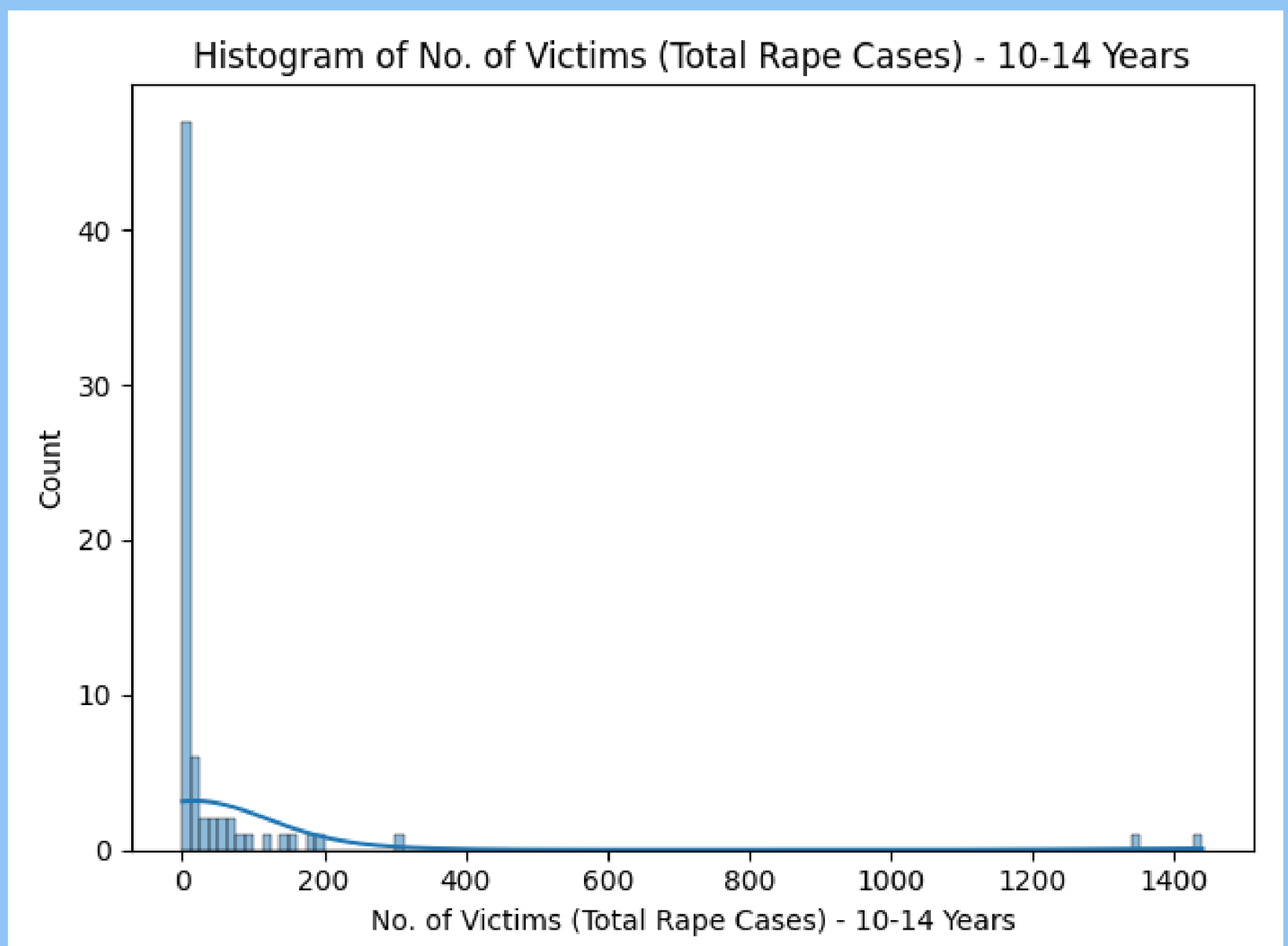

Histogram of No. of Victims (Incest Rape Cases) - 18 - 30 Years
Count
0
10
20
30
40
0
25
50
75
100
125
150
175
No. of Victims (Incest Rape Cases) - 18 - 30 Years

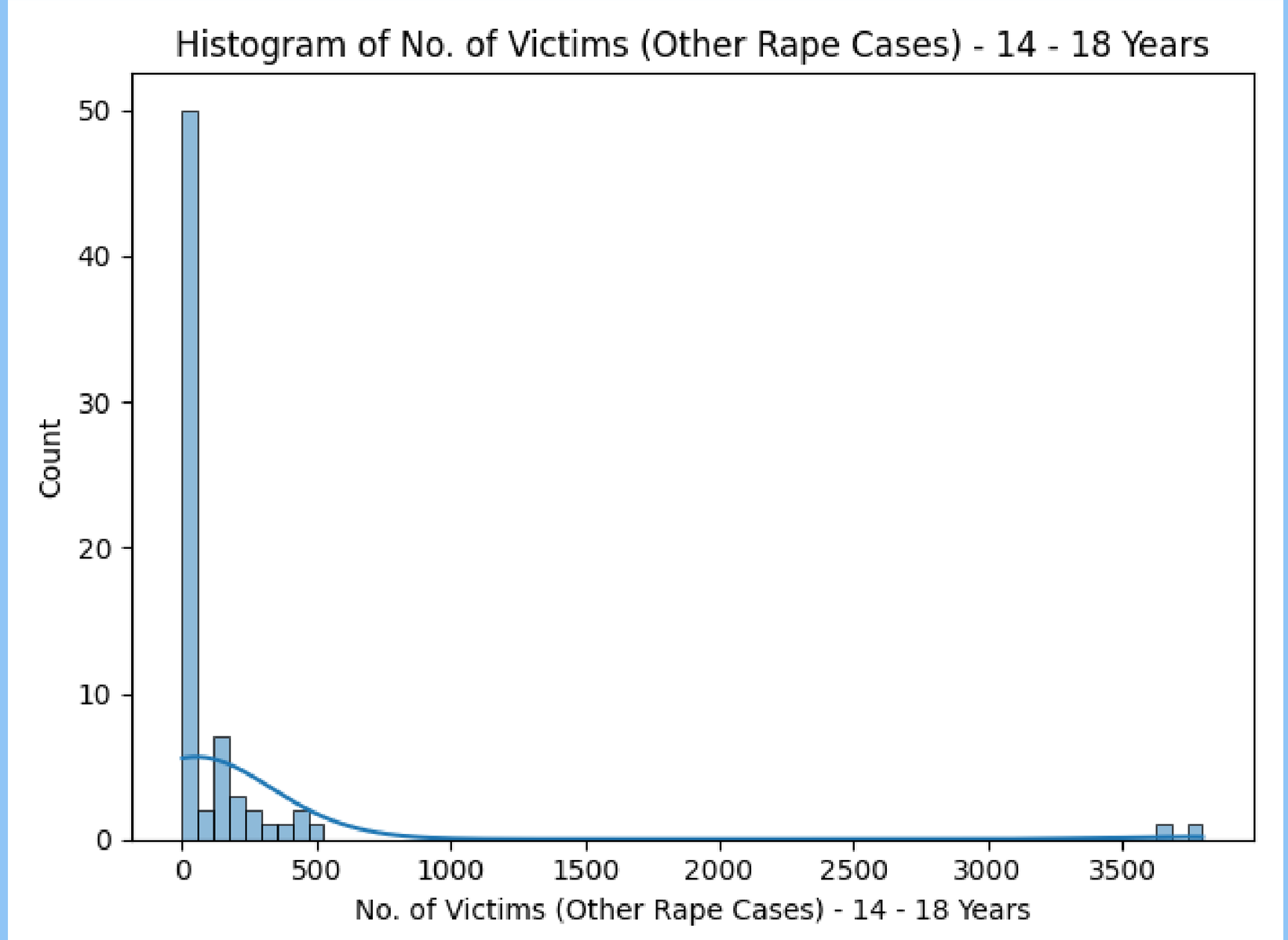

Histogram of No. of Victims (Incest Rape Cases) - 30 - 50 Years
Count
0
10
20
30
40
50
60
70
0
10
20
30
40
50
60
70
No. of Victims (Incest Rape Cases) - 30 - 50 Years

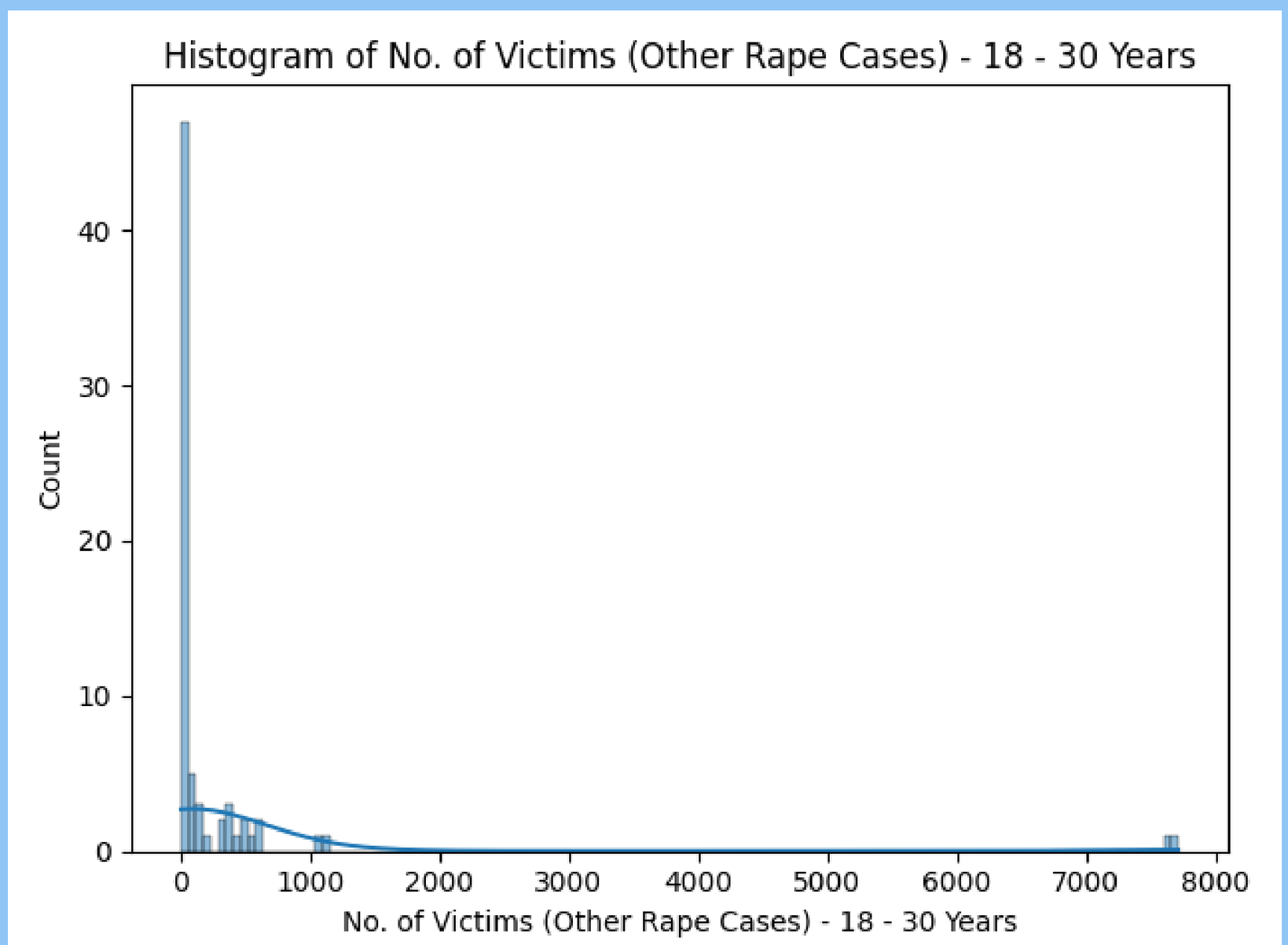

Histogram of No. of Victims (Incest Rape Cases) - Total Victims
Count
0
5
10
15
20
25
30
35
40
0
100
200
300
400
No. of Victims (Incest Rape Cases) - Total Victims

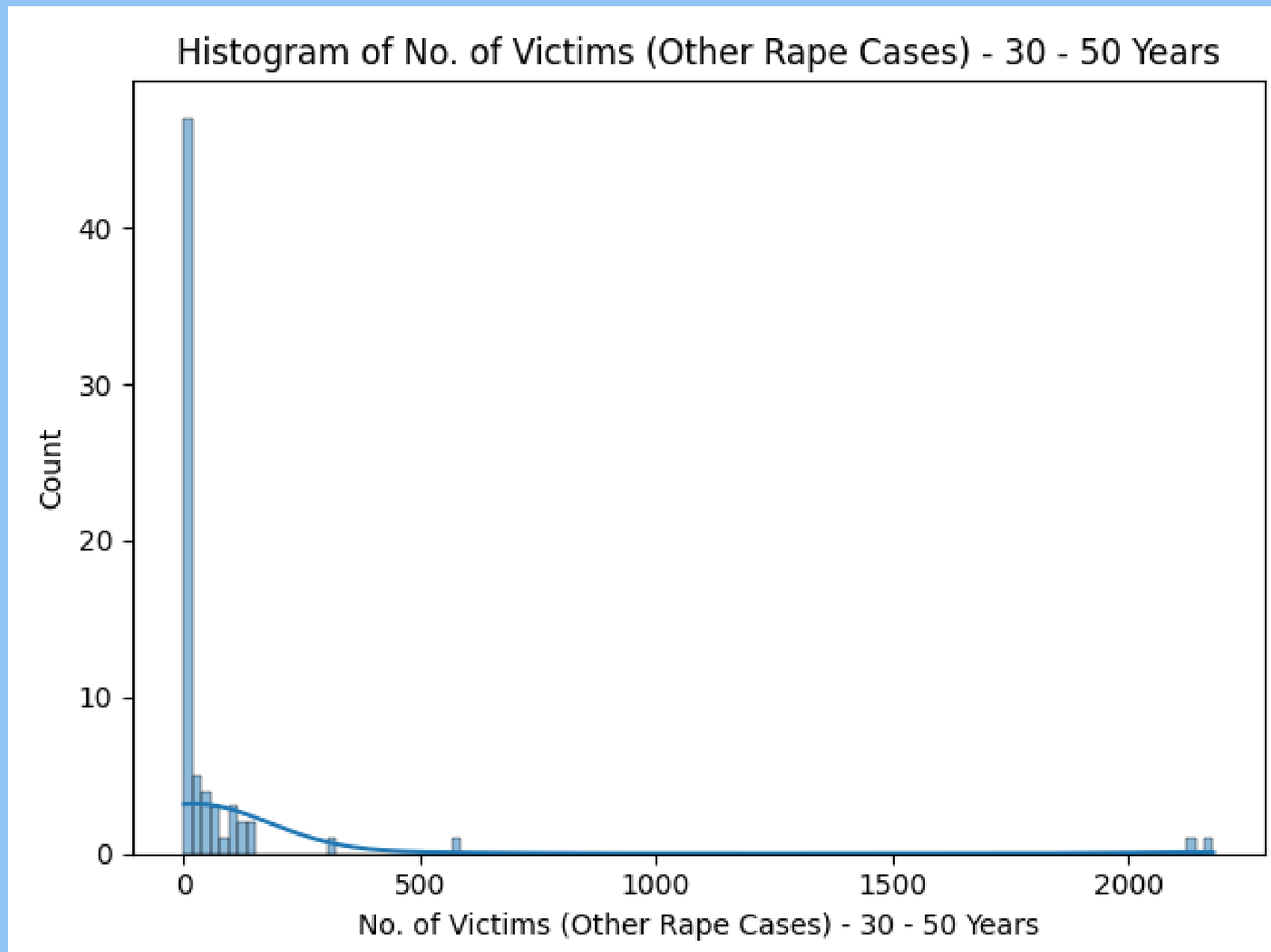

Histogram of No. of Victims (Incest Rape Cases) above 50 Years
Count
70
60
50
40
30
20
10
0
0.0
0.2
0.4
0.6
0.8
1.0
No. of Victims (Incest Rape Cases) above 50 Years

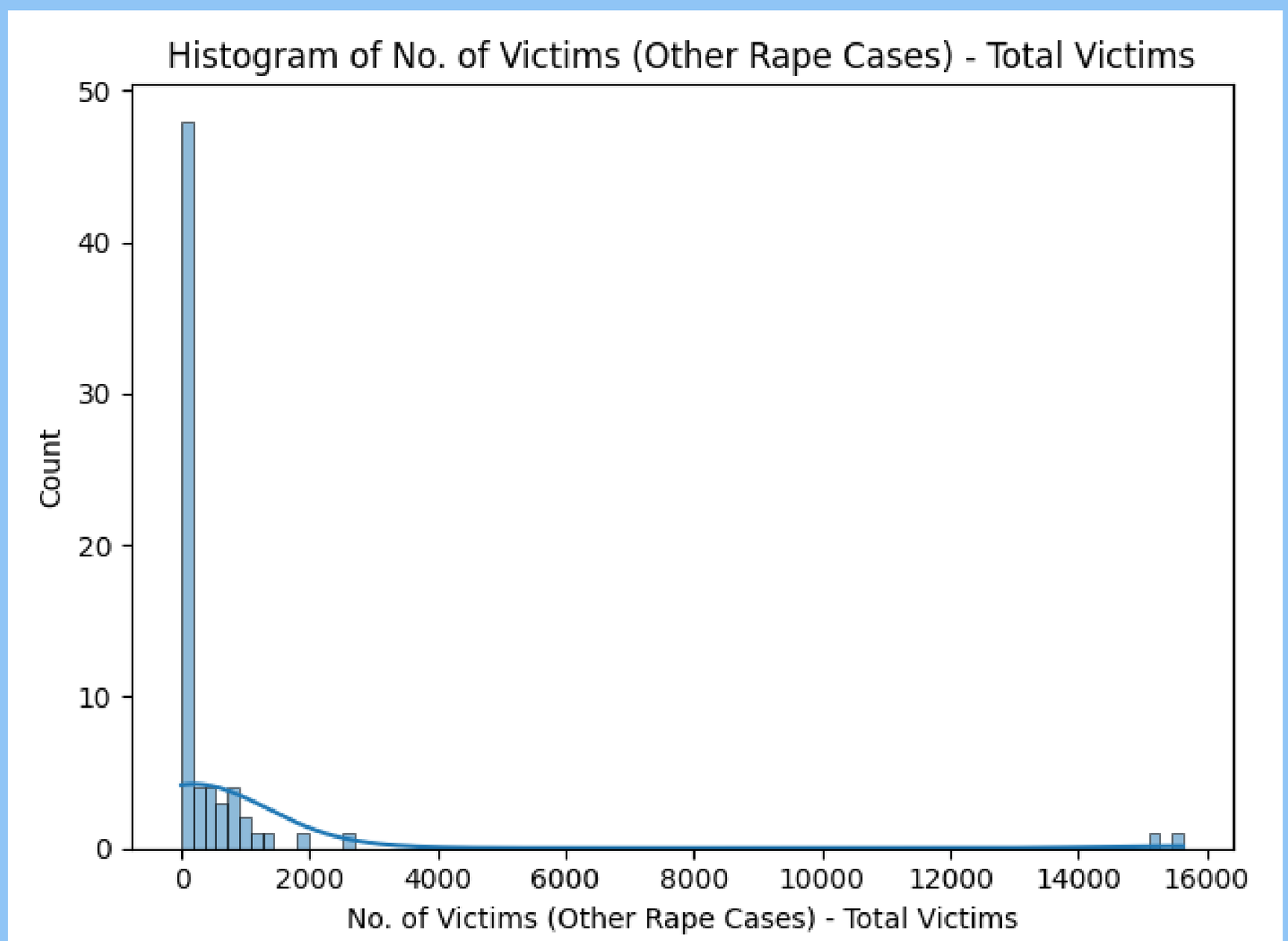

Histogram of No. of Victims (Incest Rape Cases) upto 10 Years
Count
50
40
30
20
10
0
0
5
10
15
20
25
No. of Victims (Incest Rape Cases) upto 10 Years

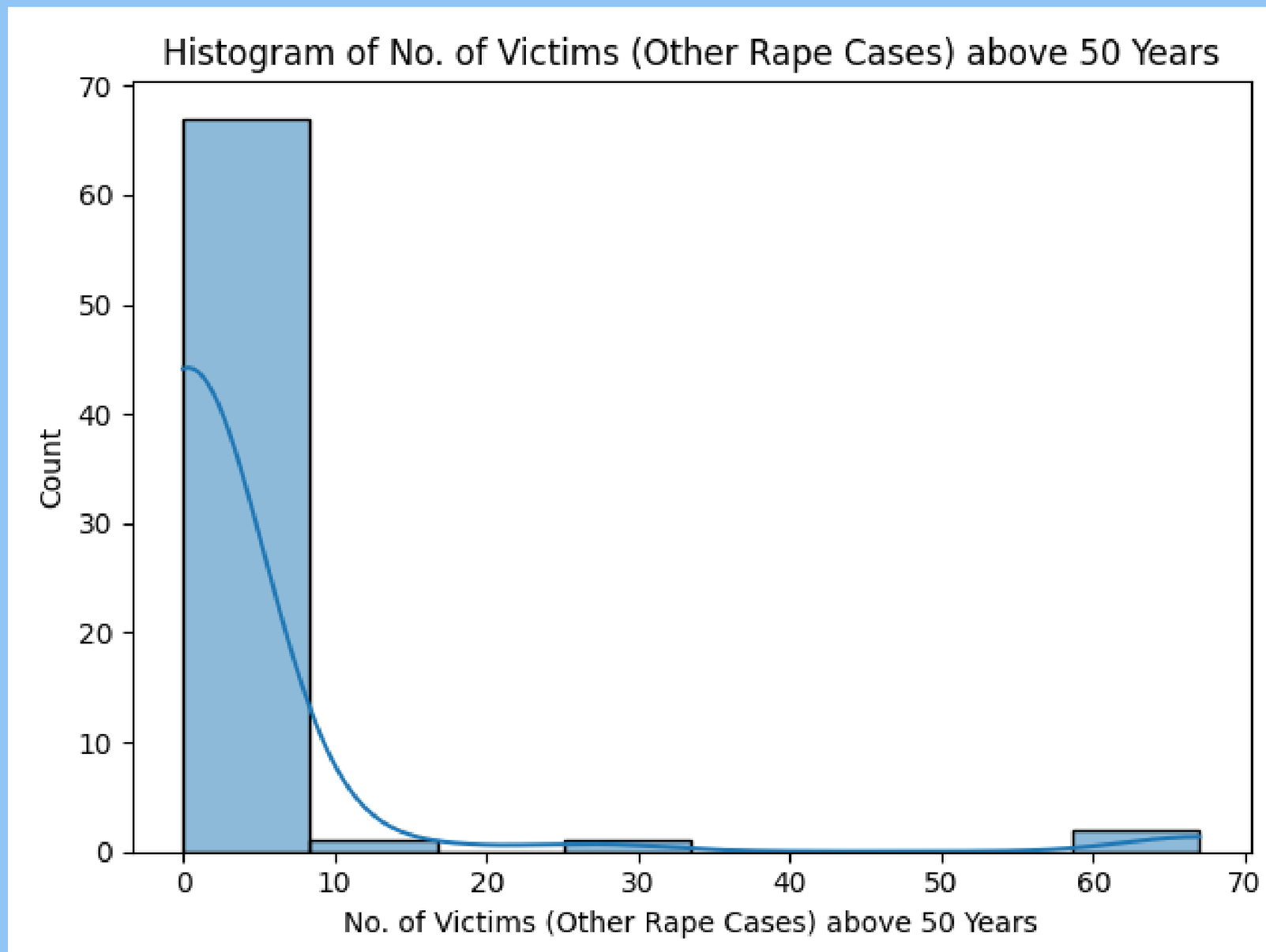

Histogram of No. of Victims (Total Rape Cases) - 10-14 Years
Count
0
10
20
30
40
0
200
400
600
800
1000
1200
1400
No. of Victims (Total Rape Cases) - 10-14 Years

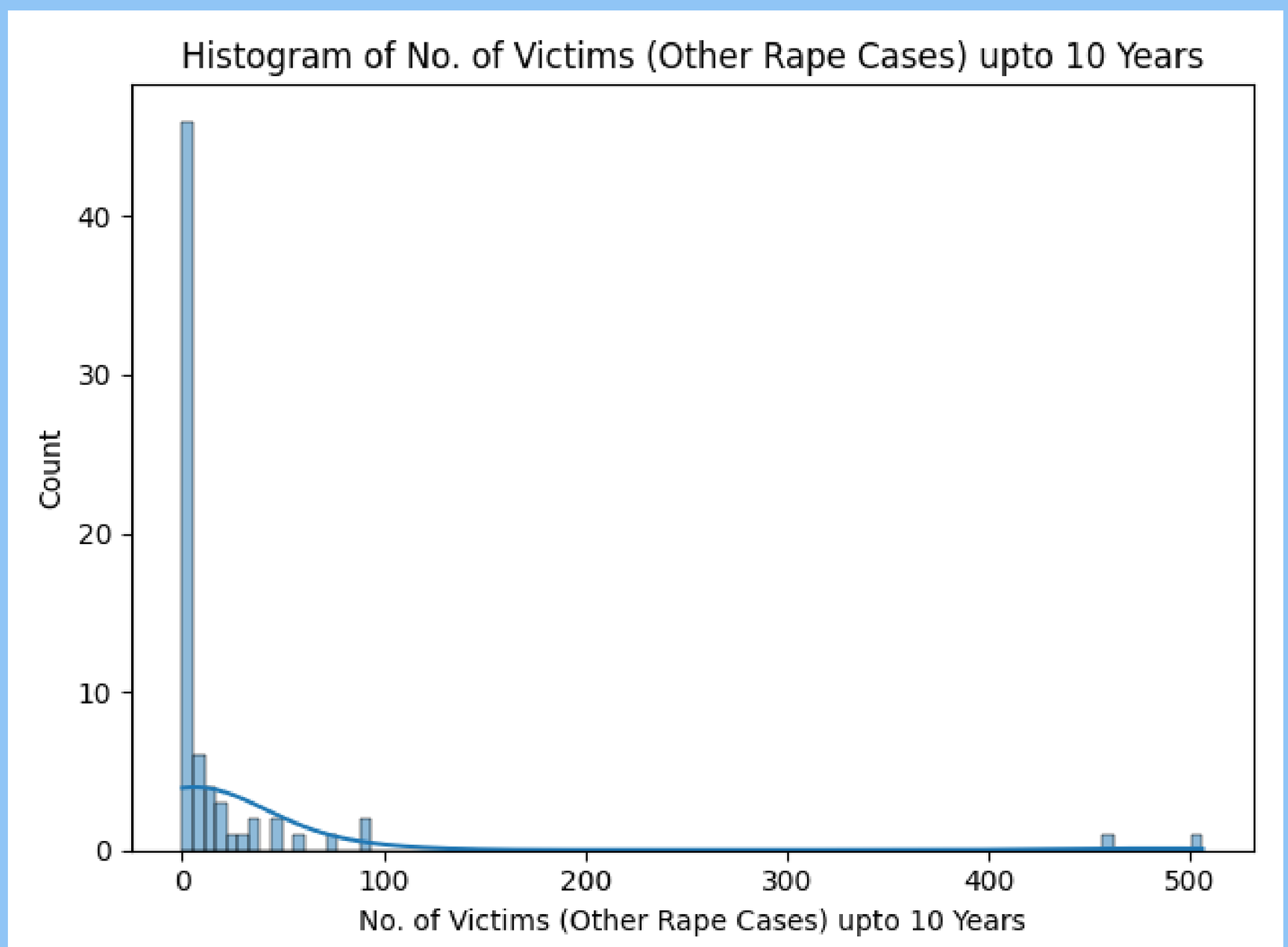

Histogram of No. of Victims (Other Rape Cases) - 14 - 18 Years
Count
0
10
20
30
40
50
0
500
1000
1500
2000
2500
3000
3500
No. of Victims (Other Rape Cases) - 14 - 18 Years

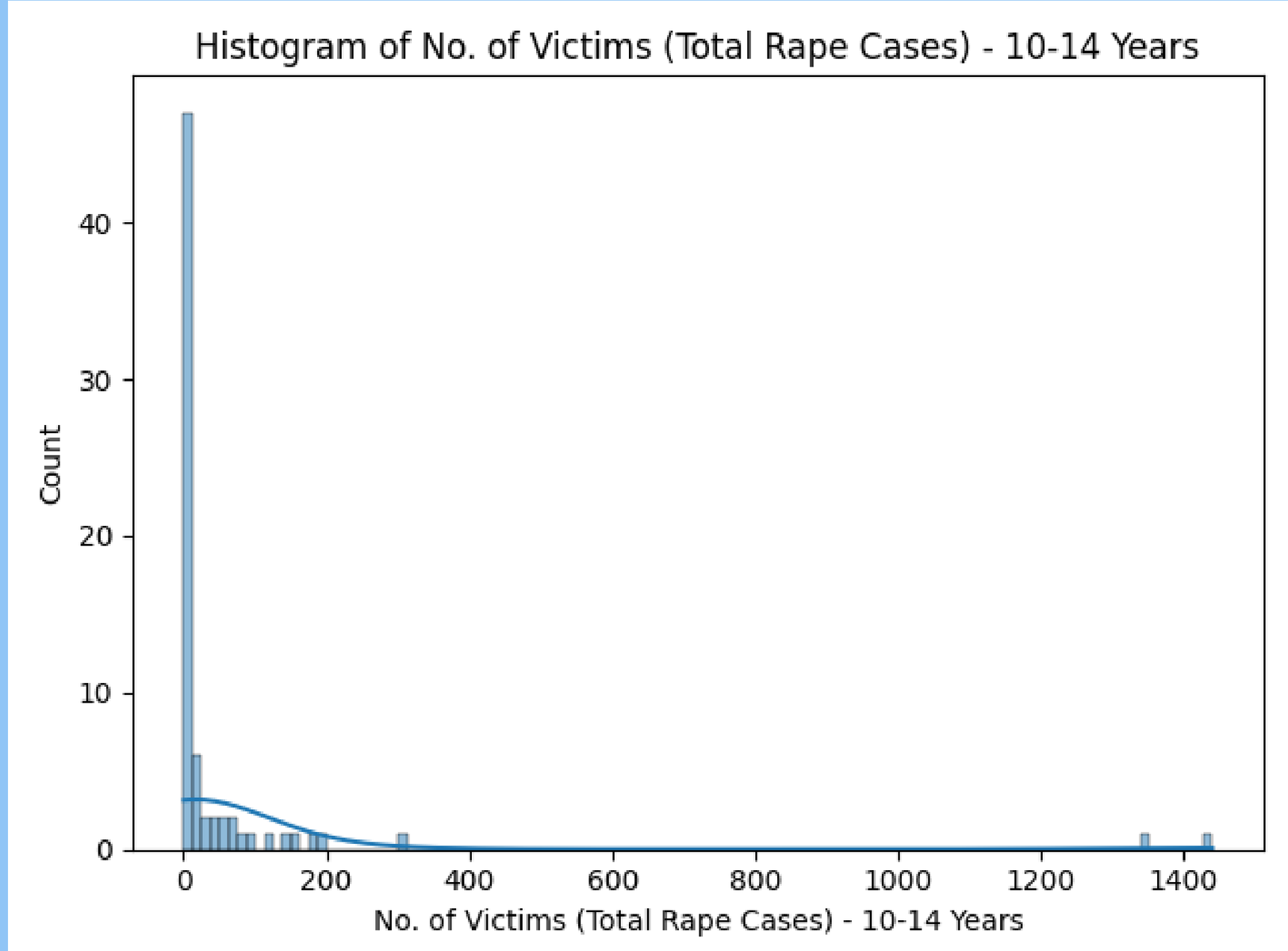

Histogram of No. of Victims (Other Rape Cases) - 18 - 30 Years
Count
0
10
20
30
40
0
1000
2000
3000
4000
5000
6000
7000
8000
No. of Victims (Other Rape Cases) - 18 - 30 Years

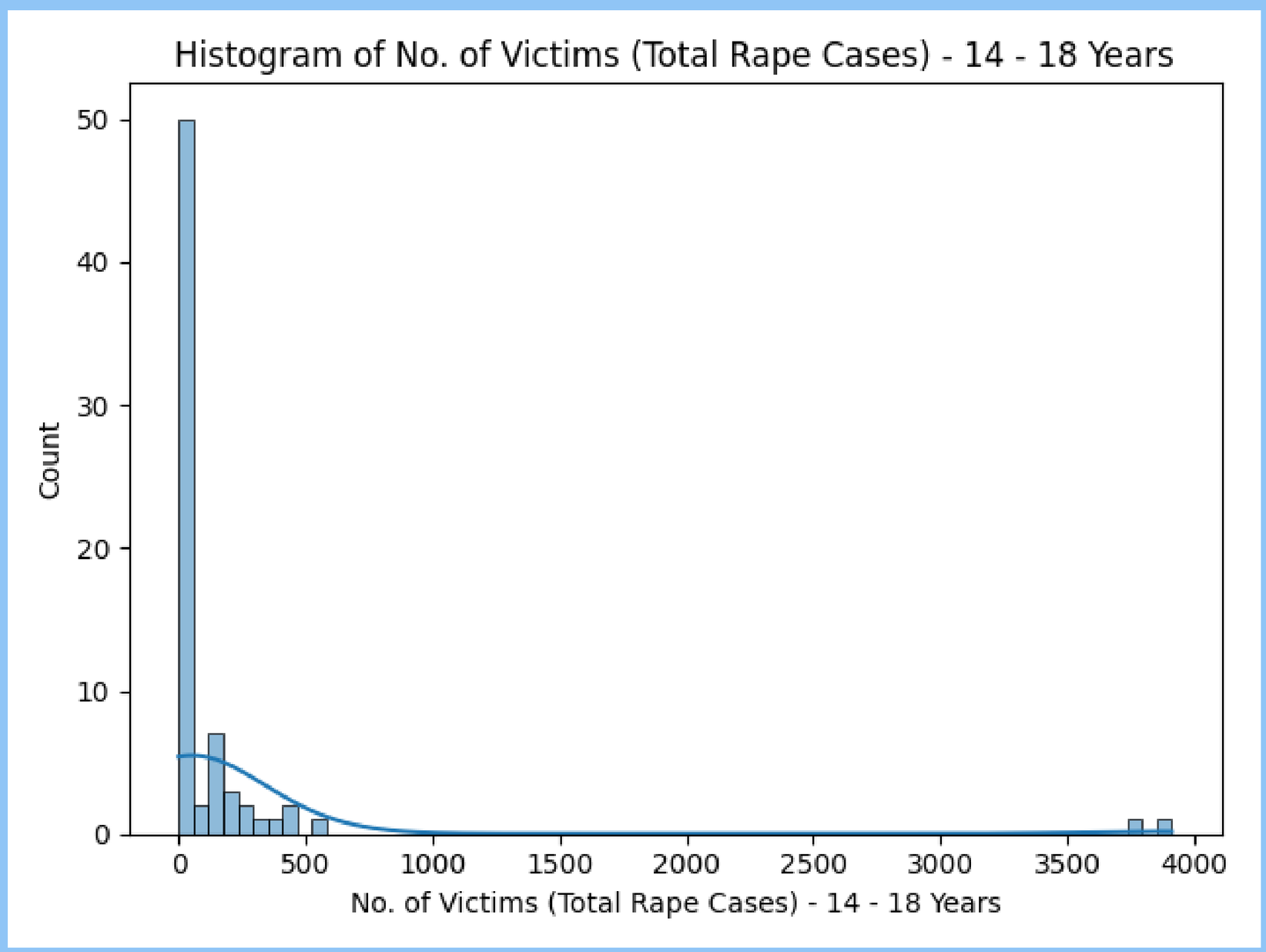

Histogram of No. of Victims (Other Rape Cases) - 30 - 50 Years
Count
0
10
20
30
40
0
500
1000
1500
2000
No. of Victims (Other Rape Cases) - 30 - 50 Years

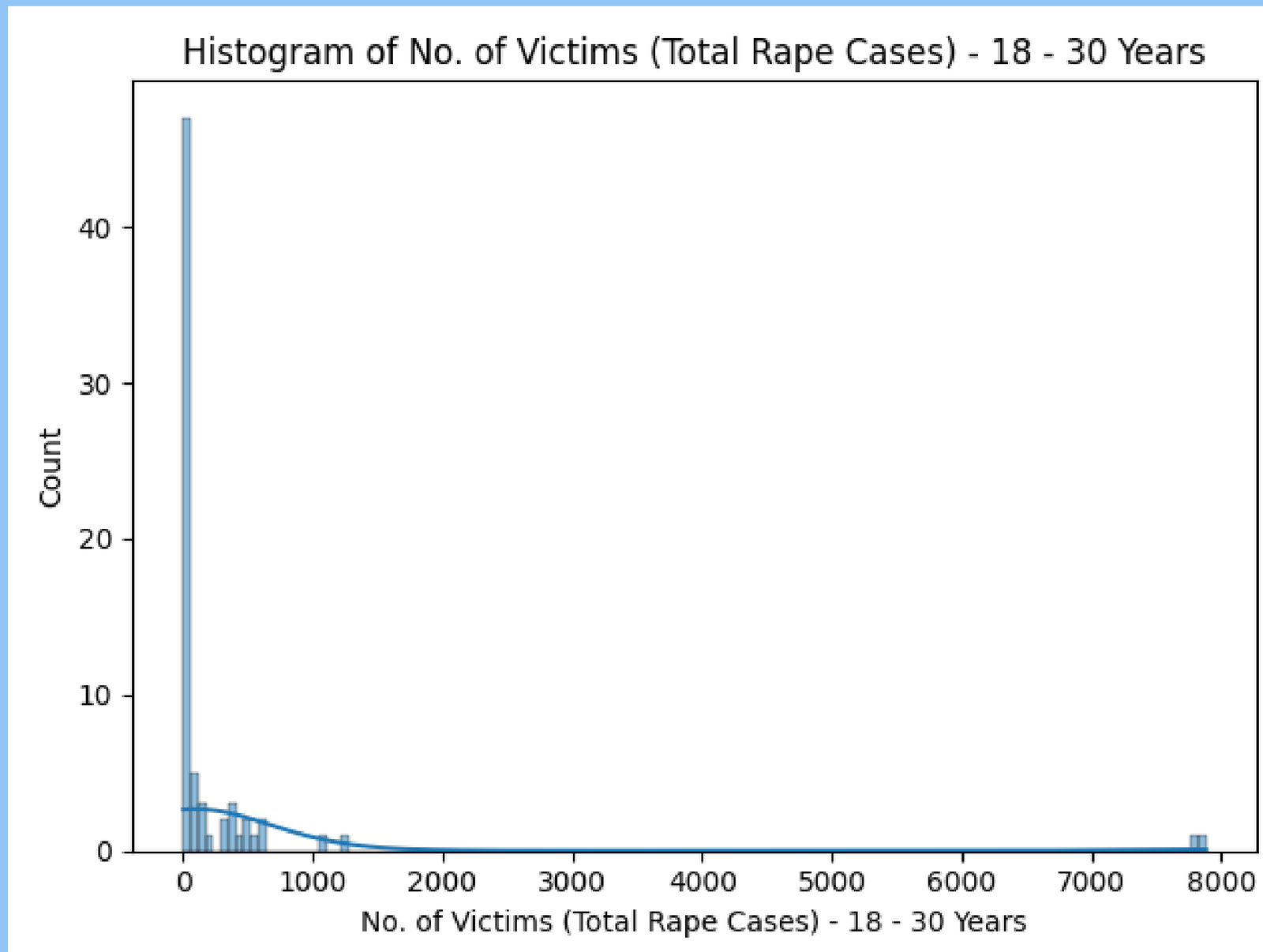

Histogram of No. of Victims (Other Rape Cases) - Total Victims
Count
0
10
20
30
40
50
0
2000
4000
6000
8000
10000
12000
14000
16000
No. of Victims (Other Rape Cases) - Total Victims

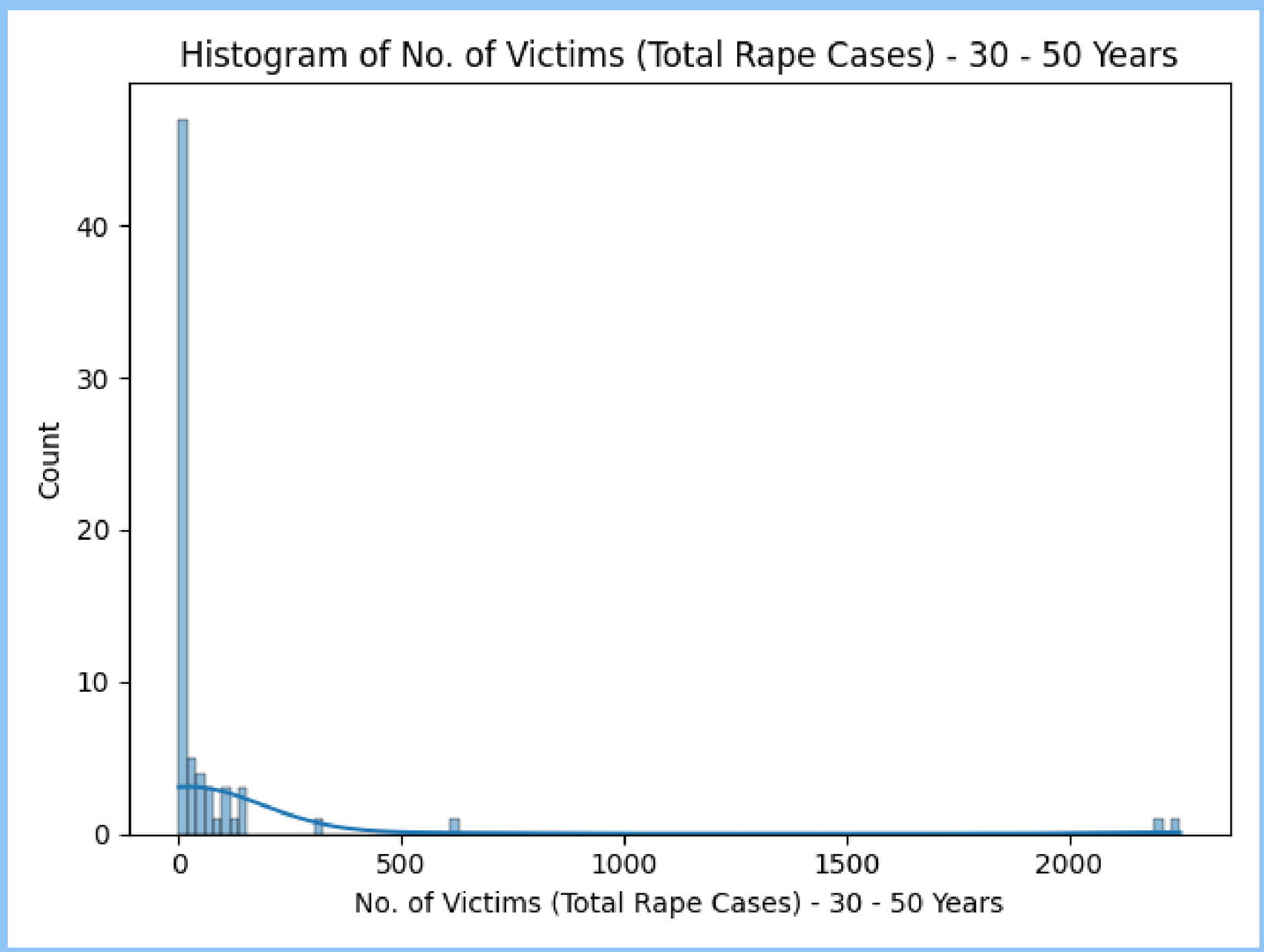

Histogram of No. of Victims (Other Rape Cases) above 50 Years
Count
0
10
20
30
40
50
60
70
0
10
20
30
40
50
60
70
No. of Victims (Other Rape Cases) above 50 Years

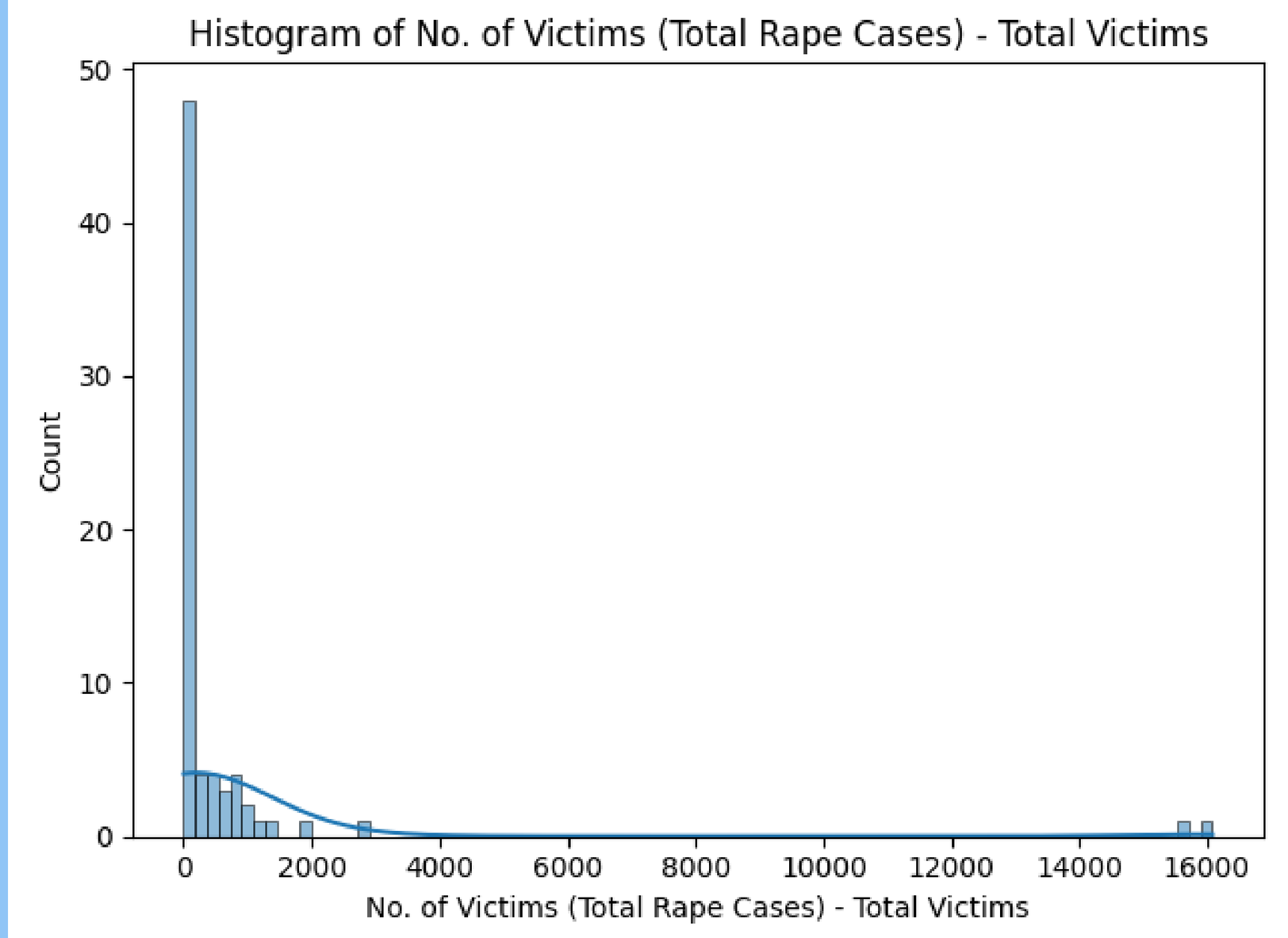

Histogram of No. of Victims (Other Rape Cases) upto 10 Years
Count
0
10
20
30
40
0
100
200
300
400
500
No. of Victims (Other Rape Cases) upto 10 Years

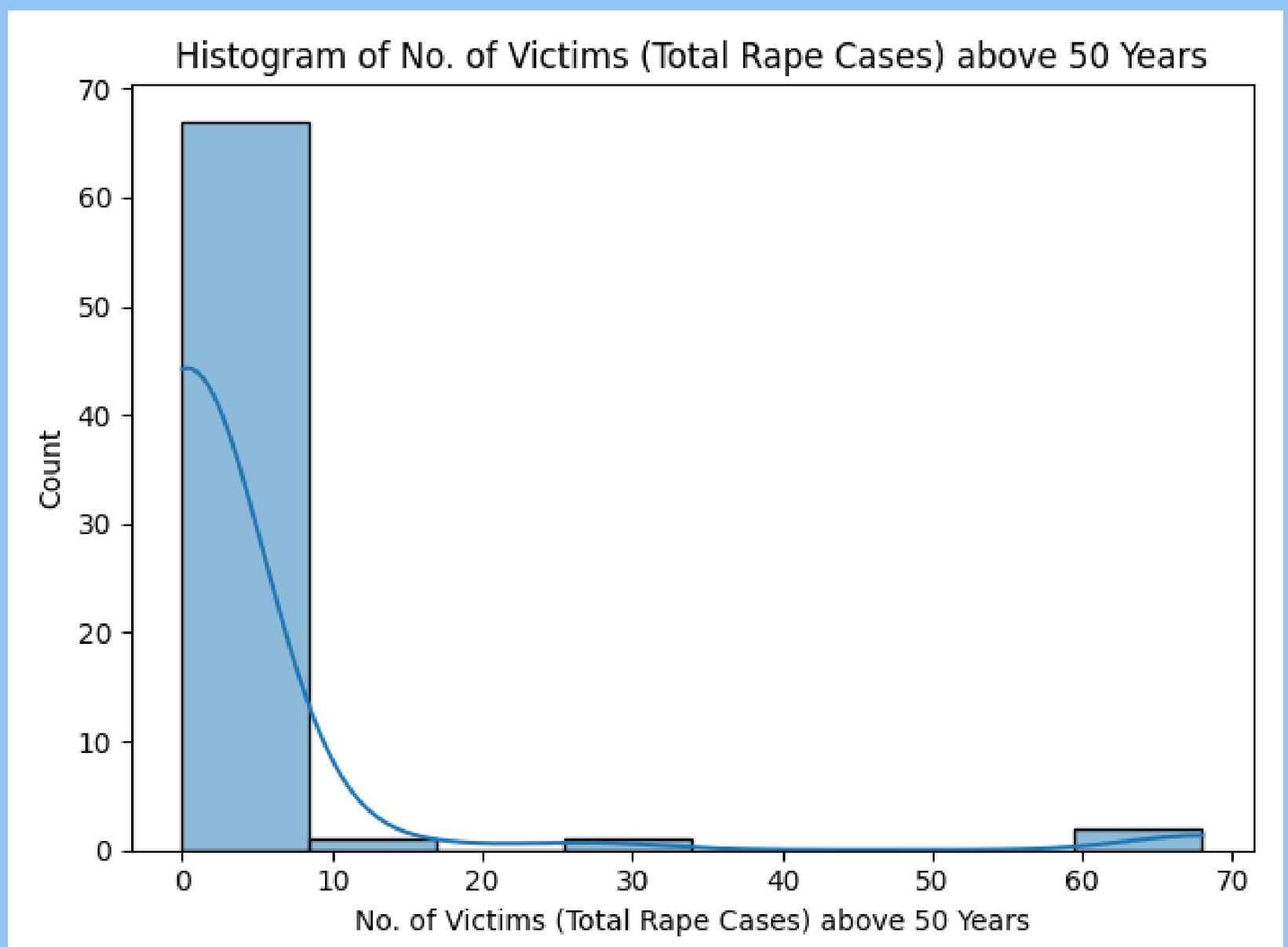

Histogram of No. of Victims (Total Rape Cases) - 10-14 Years
Count
0
10
20
30
40
0
200
400
600
800
1000
1200
1400
No. of Victims (Total Rape Cases) - 10-14 Years

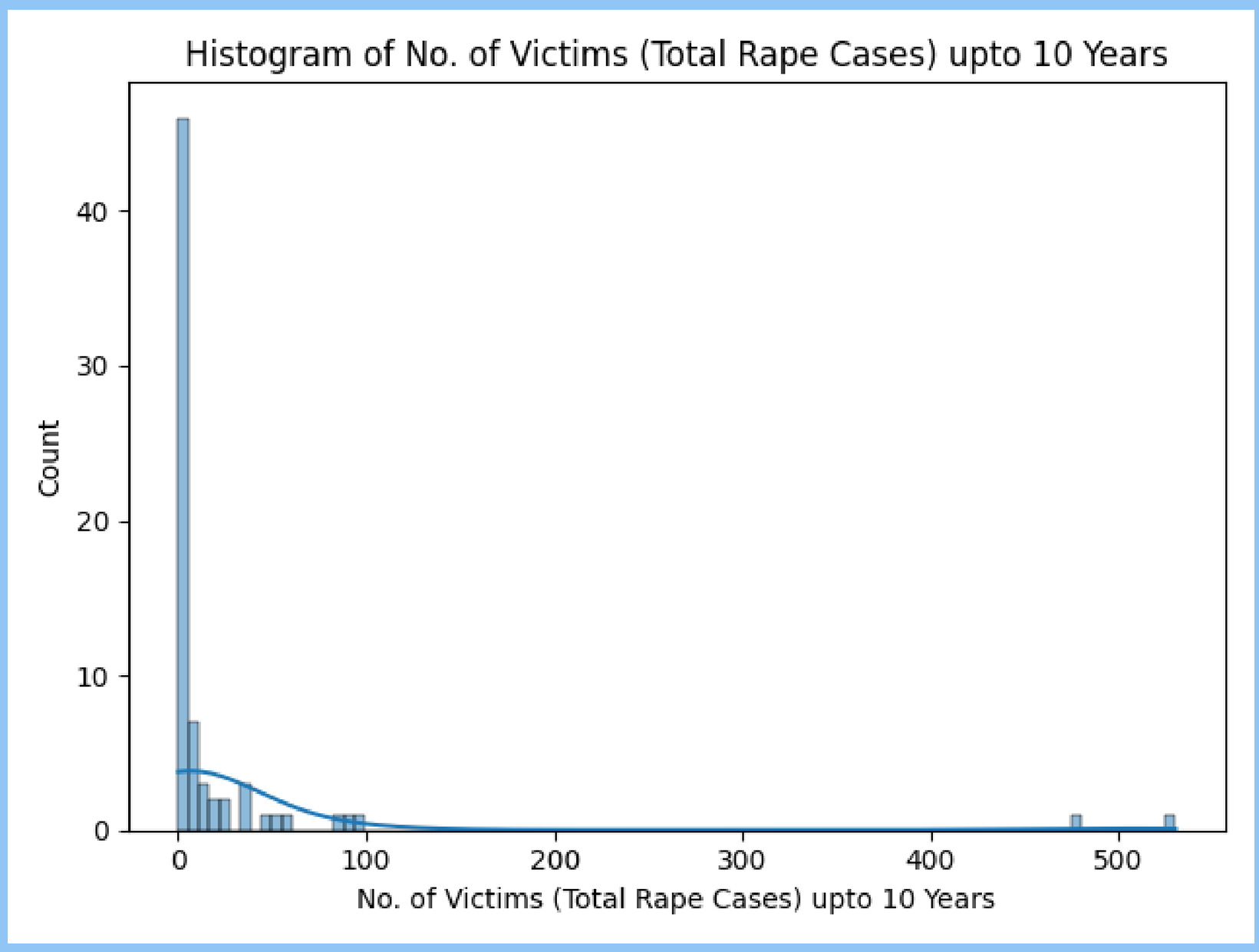

Histogram of No. of Victims (Total Rape Cases) - 14 - 18 Years
Count
50
40
30
20
10
0
0
500
1000
1500
2000
2500
3000
3500
4000
No. of Victims (Total Rape Cases) - 14 - 18 Years


Histogram of No. of Victims (Total Rape Cases) - 18 - 30 Years
Count
40
30
20
10
0
0
1000
2000
3000
4000
5000
6000
7000
8000
No. of Victims (Total Rape Cases) - 18 - 30 Years

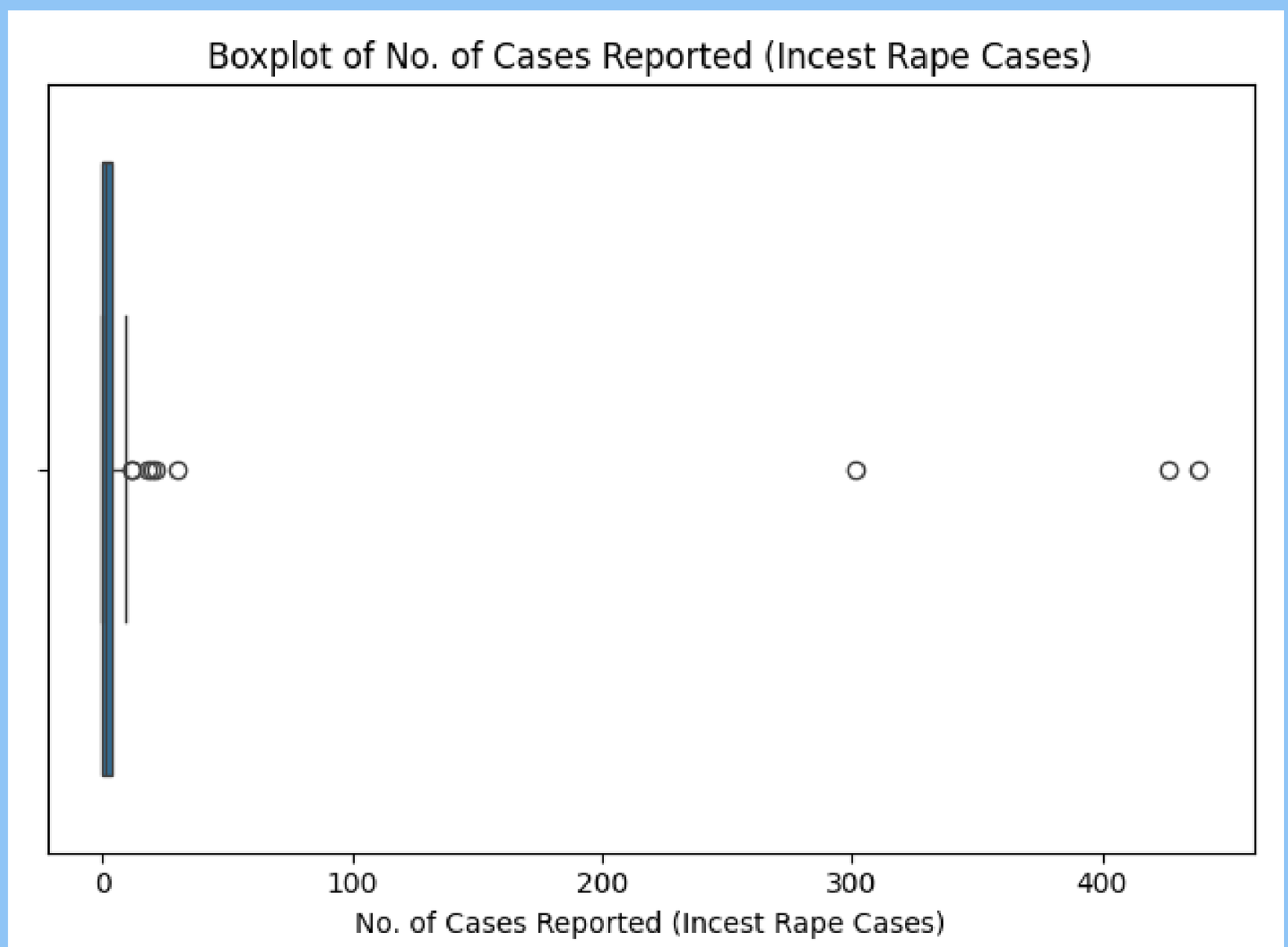

Histogram of No. of Victims (Total Rape Cases) - 30 - 50 Years
Count
0
10
20
30
40
0
500
1000
1500
2000
No. of Victims (Total Rape Cases) - 30 - 50 Years

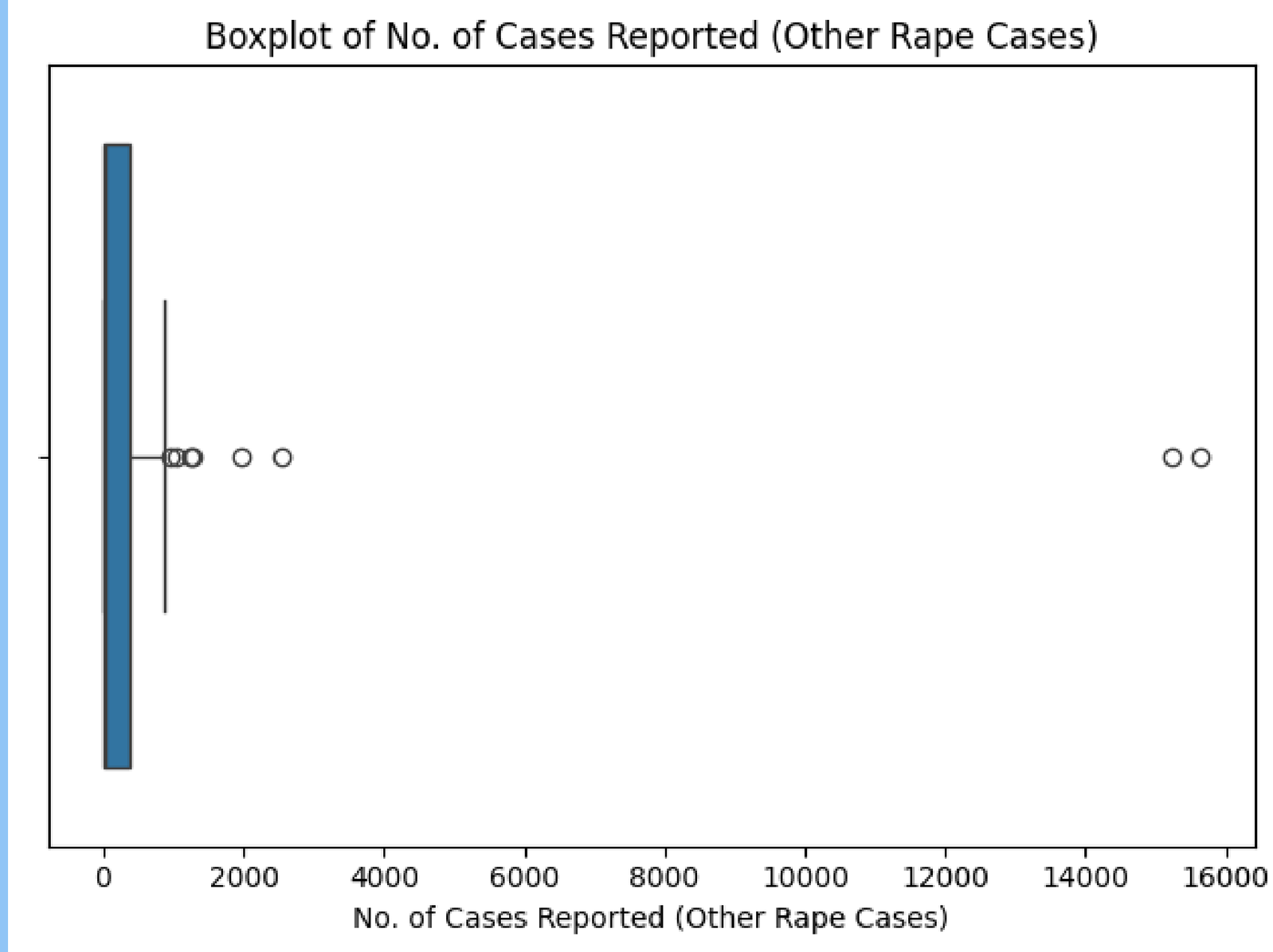

Histogram of No. of Victims (Total Rape Cases) - Total Victims
Count
0
10
20
30
40
50
0
2000
4000
6000
8000
10000
12000
14000
16000
No. of Victims (Total Rape Cases) - Total Victims

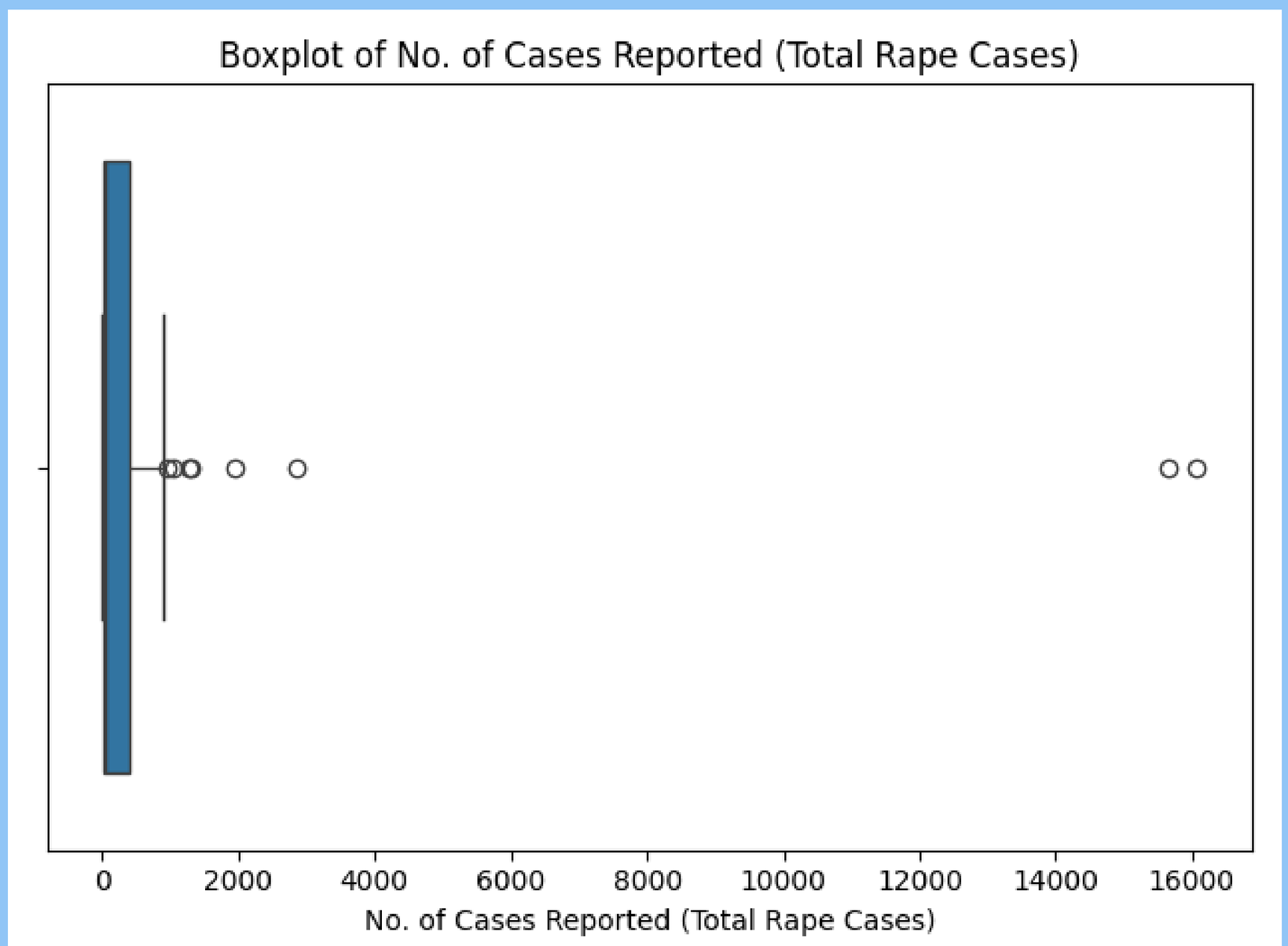

Histogram of No. of Victims (Total Rape Cases) above 50 Years
Count
70
60
50
40
30
20
10
0
0
10
20
30
40
50
60
70
No. of Victims (Total Rape Cases) above 50 Years

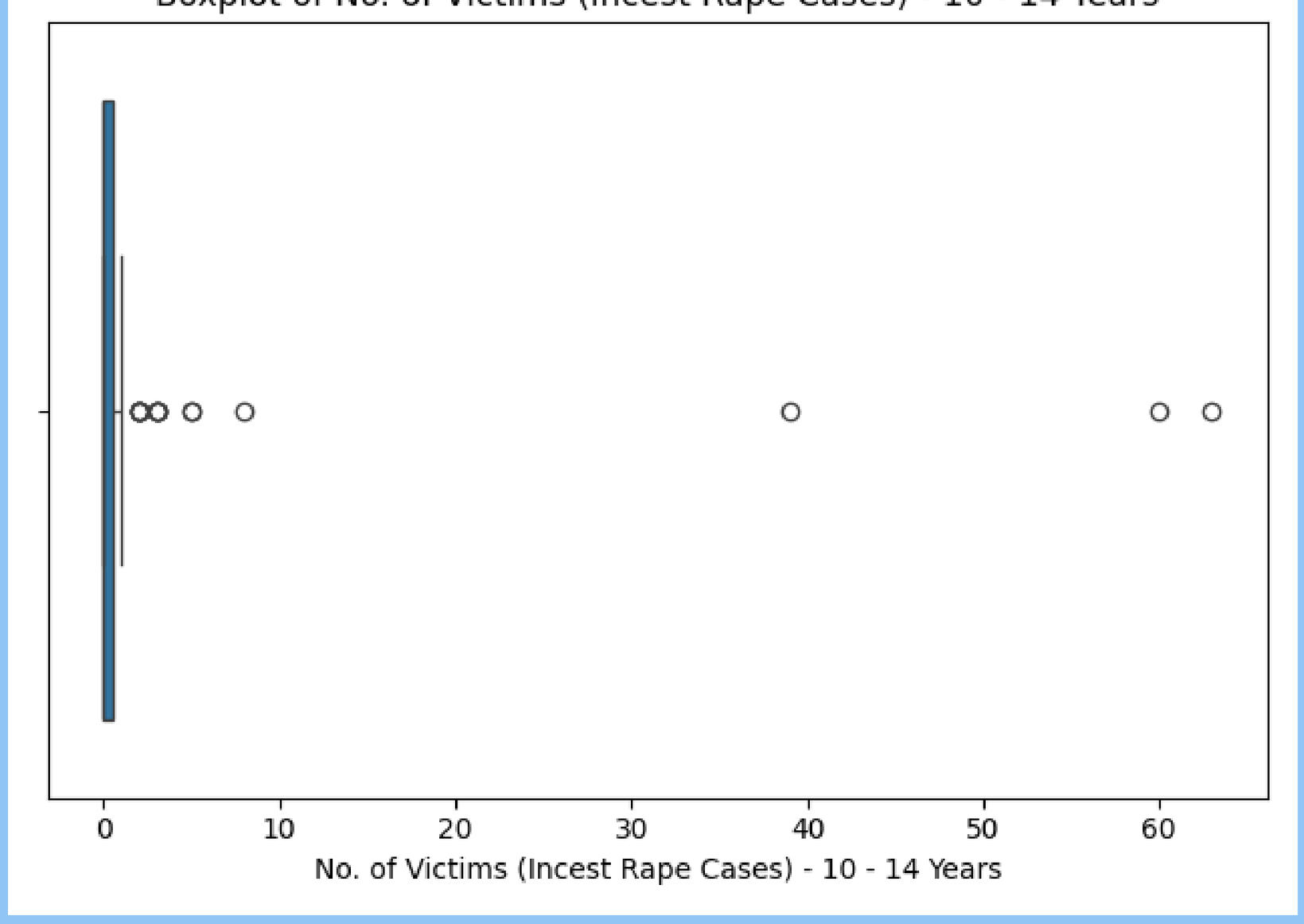

Histogram of No. of Victims (Total Rape Cases) upto 10 Years
Count
40
30
20
10
0
0
100
200
300
400
500
No. of Victims (Total Rape Cases) upto 10 Years

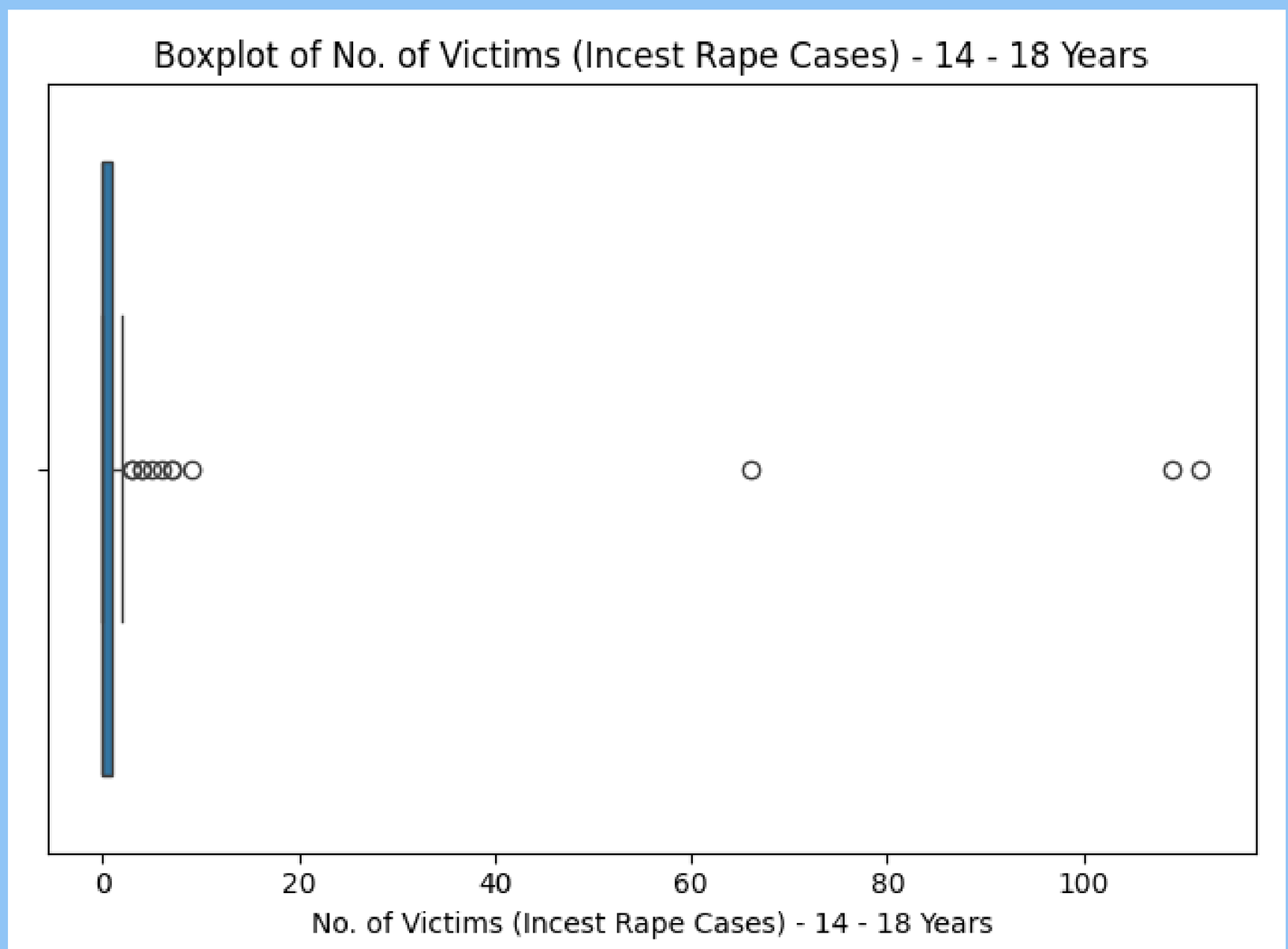
BOX PLOTS

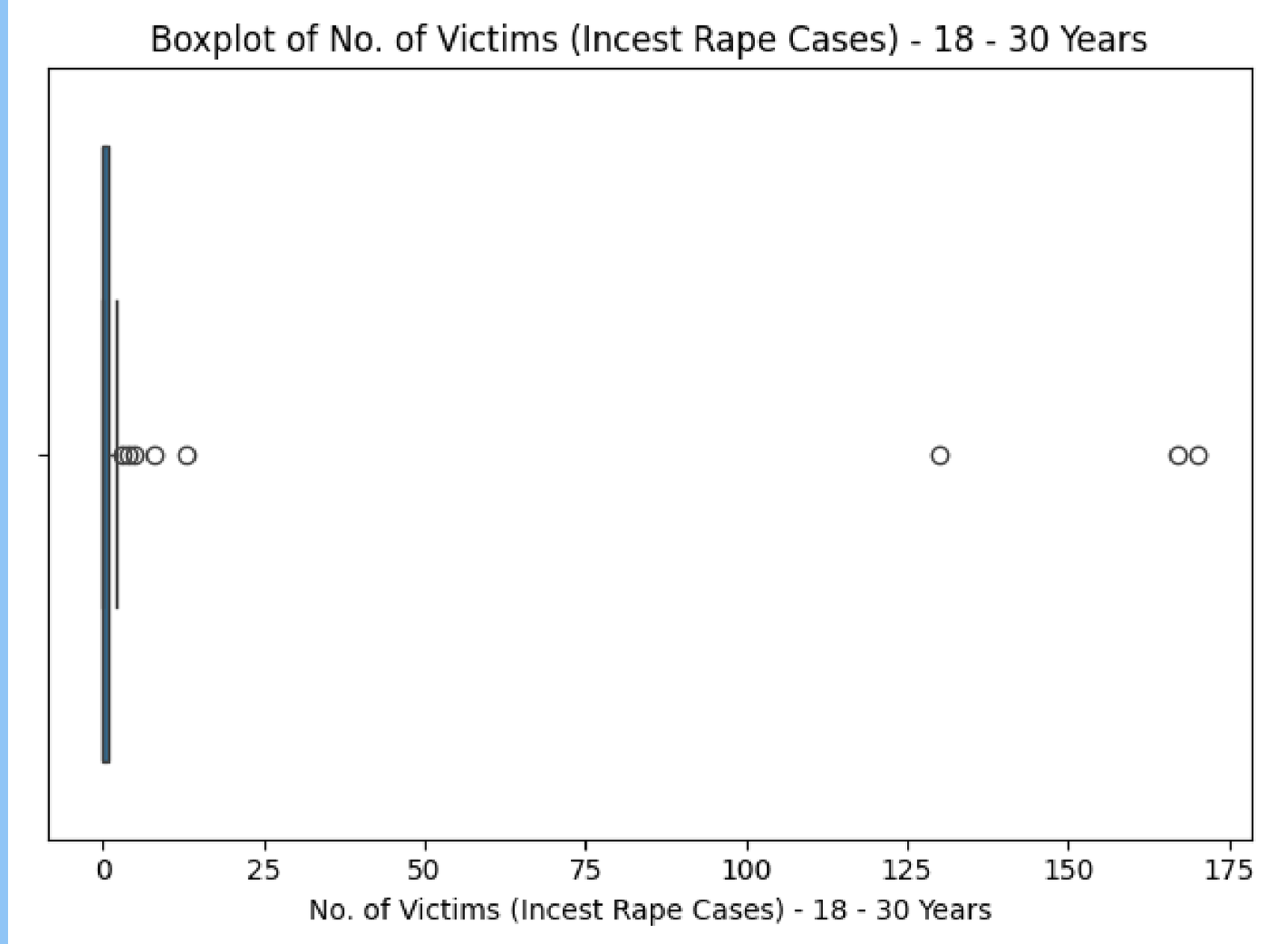

Boxplot of No. of Cases Reported (Incest Rape Cases)
0
100
200
300
400
No. of Cases Reported (Incest Rape Cases)

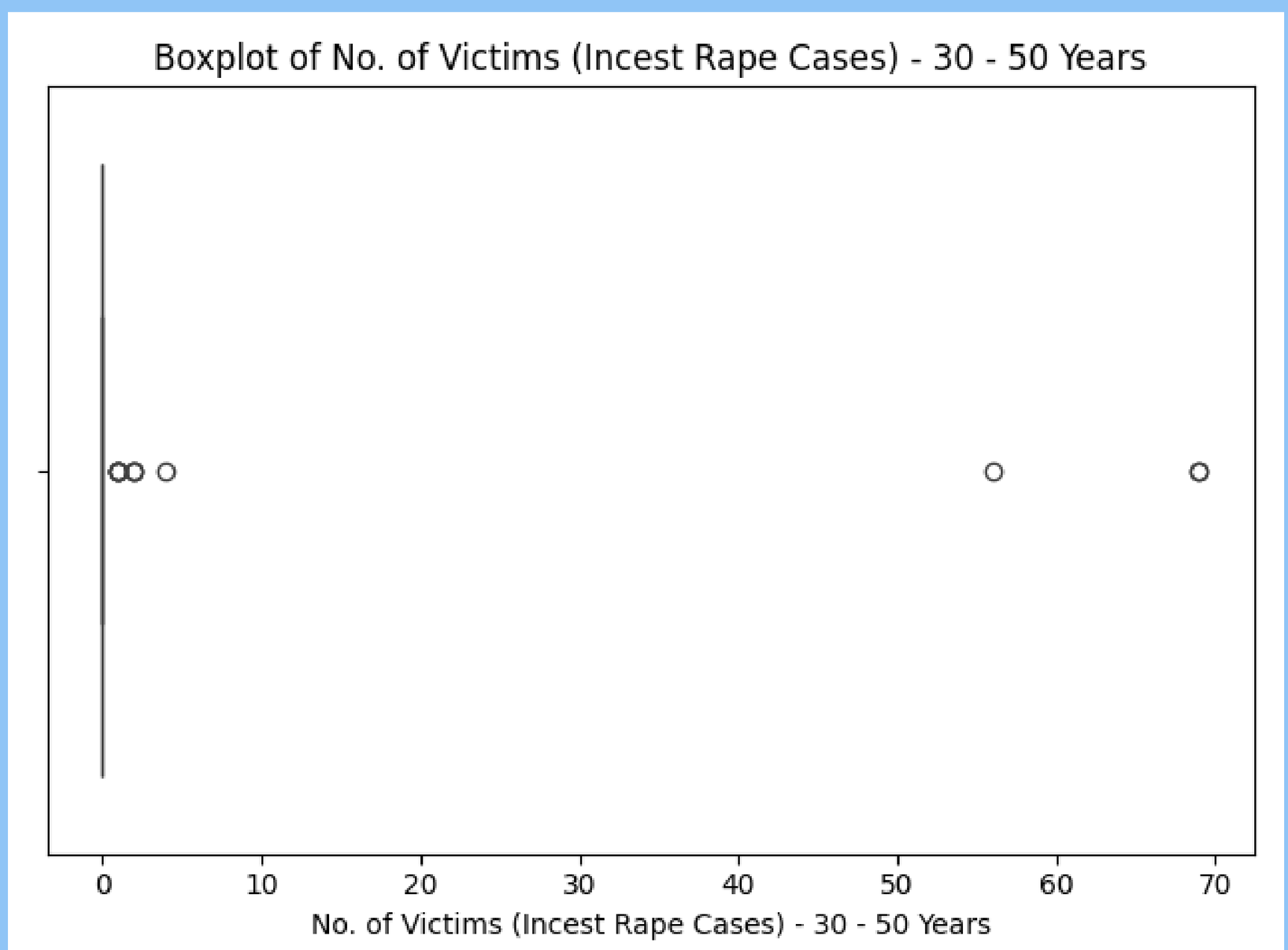

Boxplot of No. of Cases Reported (Other Rape Cases)
0
2000
4000
6000
8000
10000
12000
14000
16000
No. of Cases Reported (Other Rape Cases)

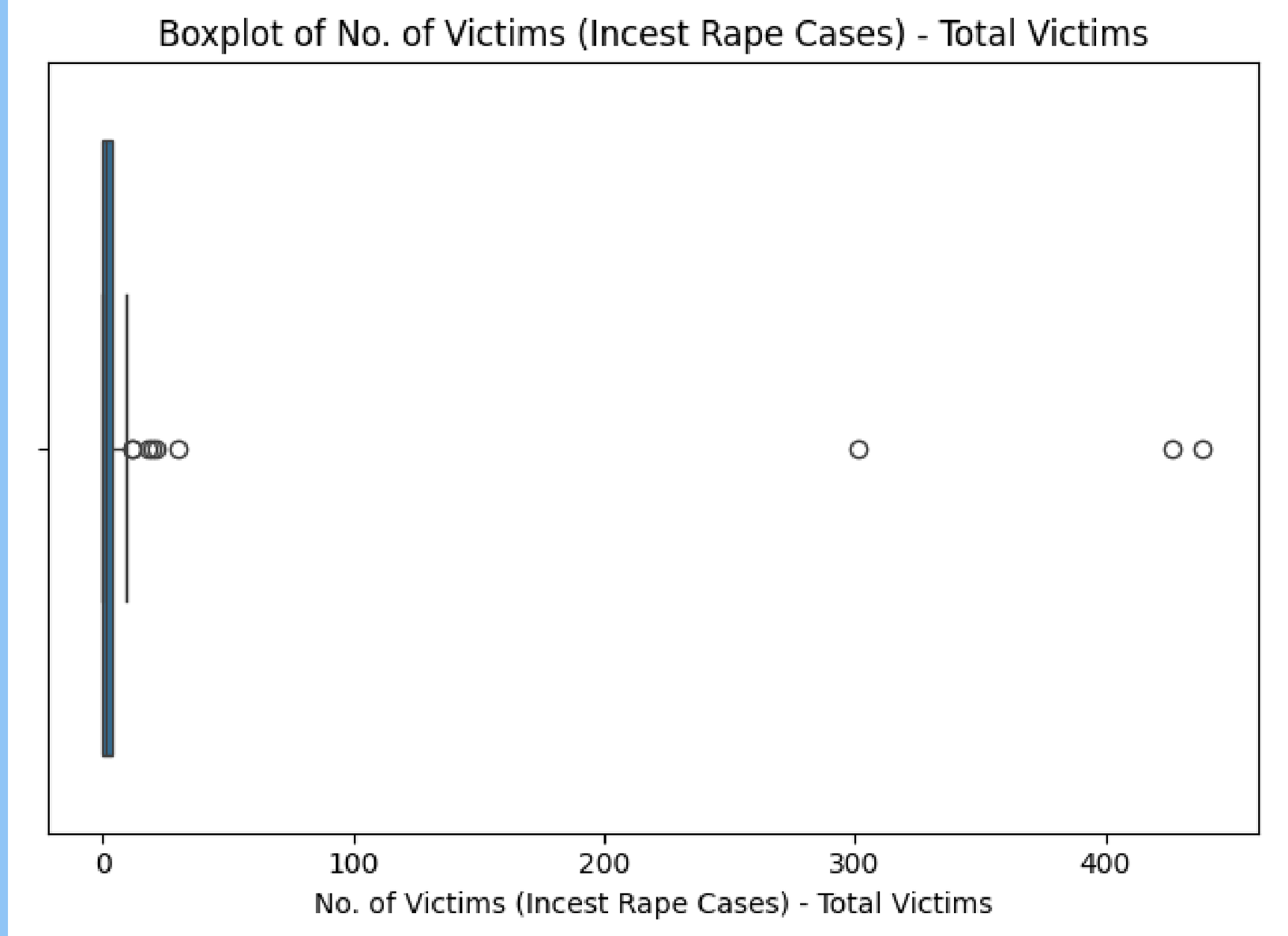

Boxplot of No. of Cases Reported (Total Rape Cases)
0
2000
4000
6000
8000
10000
12000
14000
16000
No. of Cases Reported (Total Rape Cases)

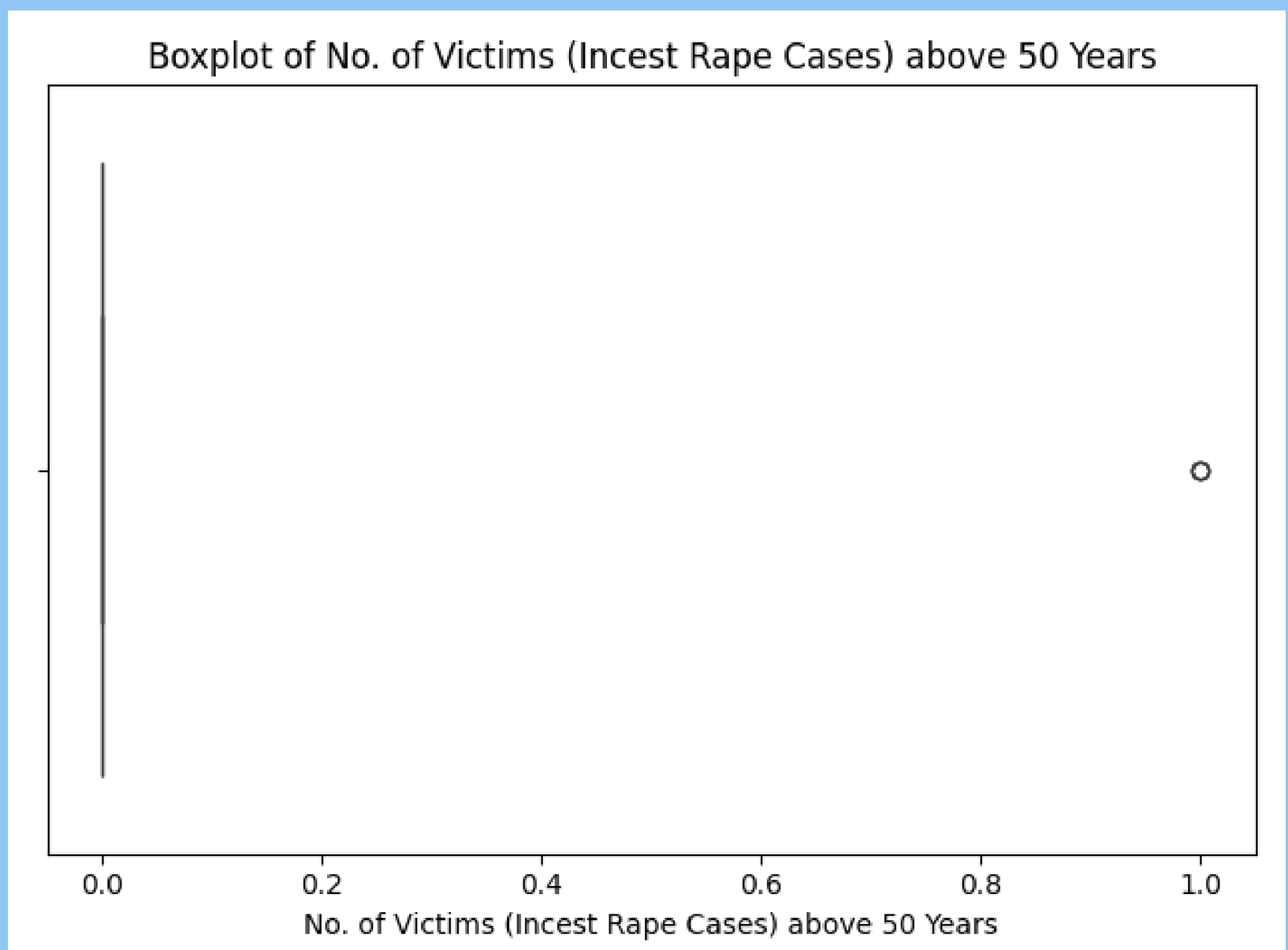

Boxplot of No. of Victims (Incest Rape Cases) - 10 - 14 Years
0
10
20
30
40
50
60
No. of Victims (Incest Rape Cases) - 10 - 14 Years

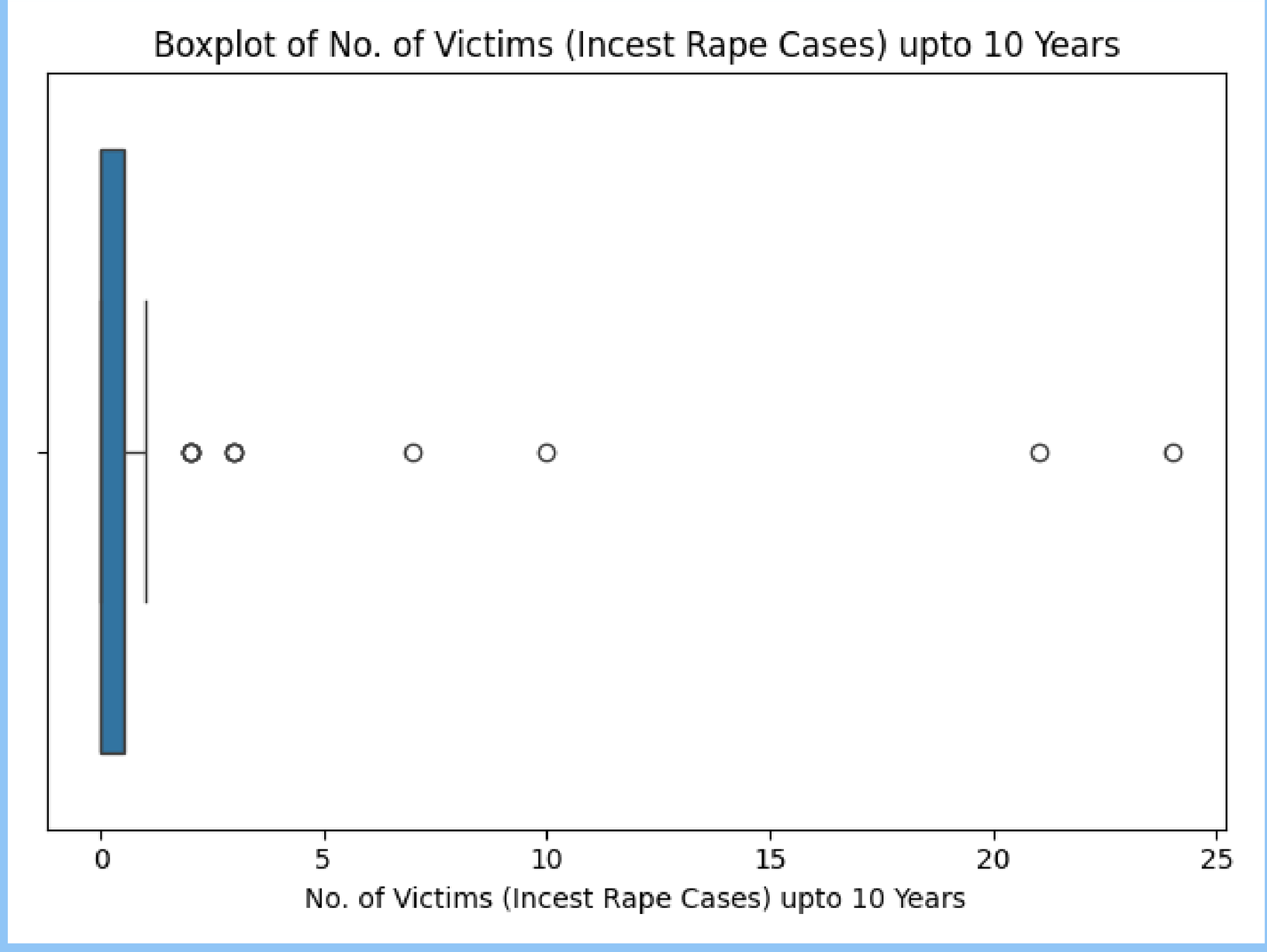

Boxplot of No. of Victims (Incest Rape Cases) - 14 - 18 Years
0
20
40
60
80
100
No. of Victims (Incest Rape Cases) - 14 - 18 Years

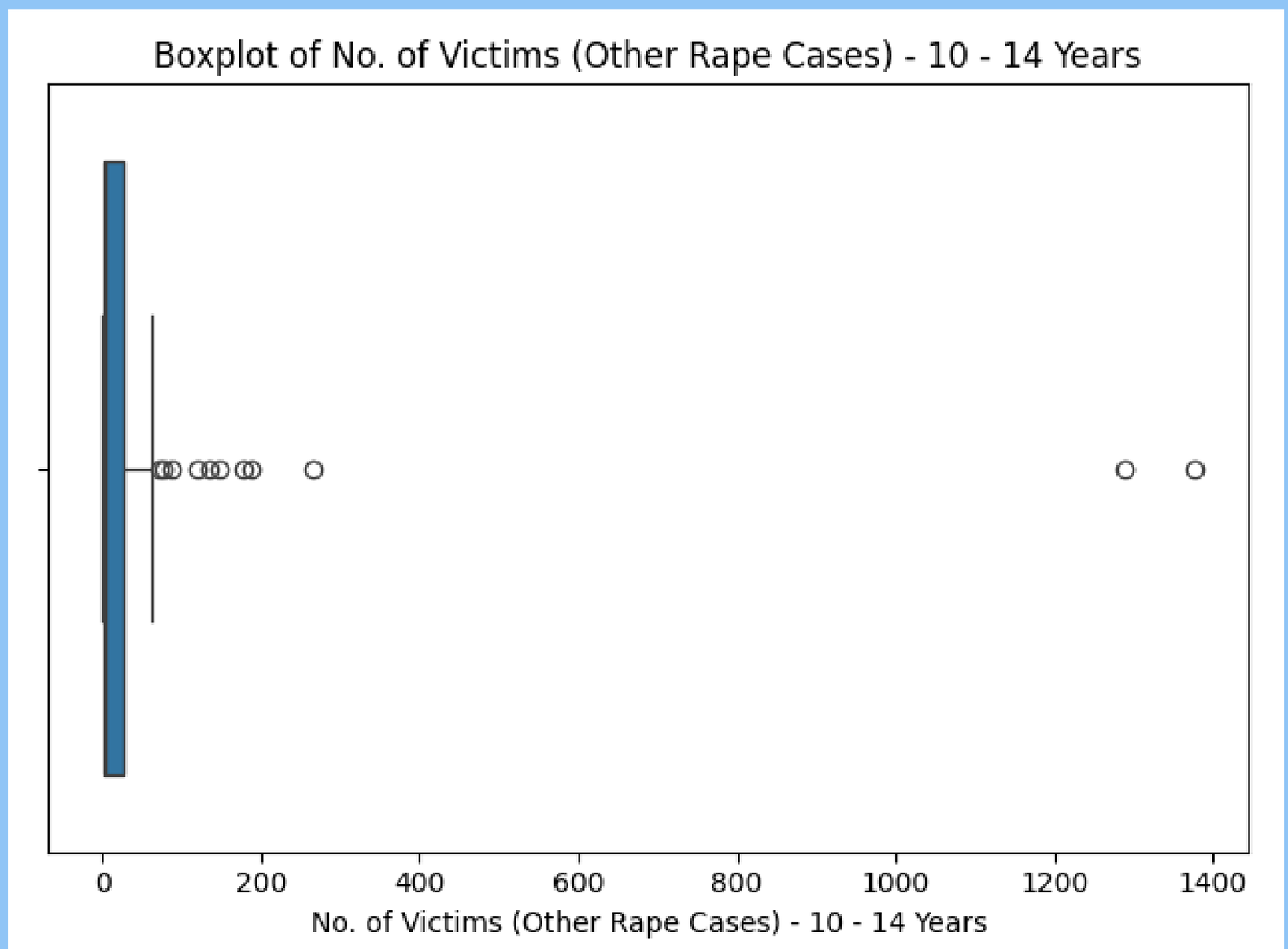

Boxplot of No. of Victims (Incest Rape Cases) - 18 - 30 Years
0
25
50
75
100
125
150
175
No. of Victims (Incest Rape Cases) - 18 - 30 Years

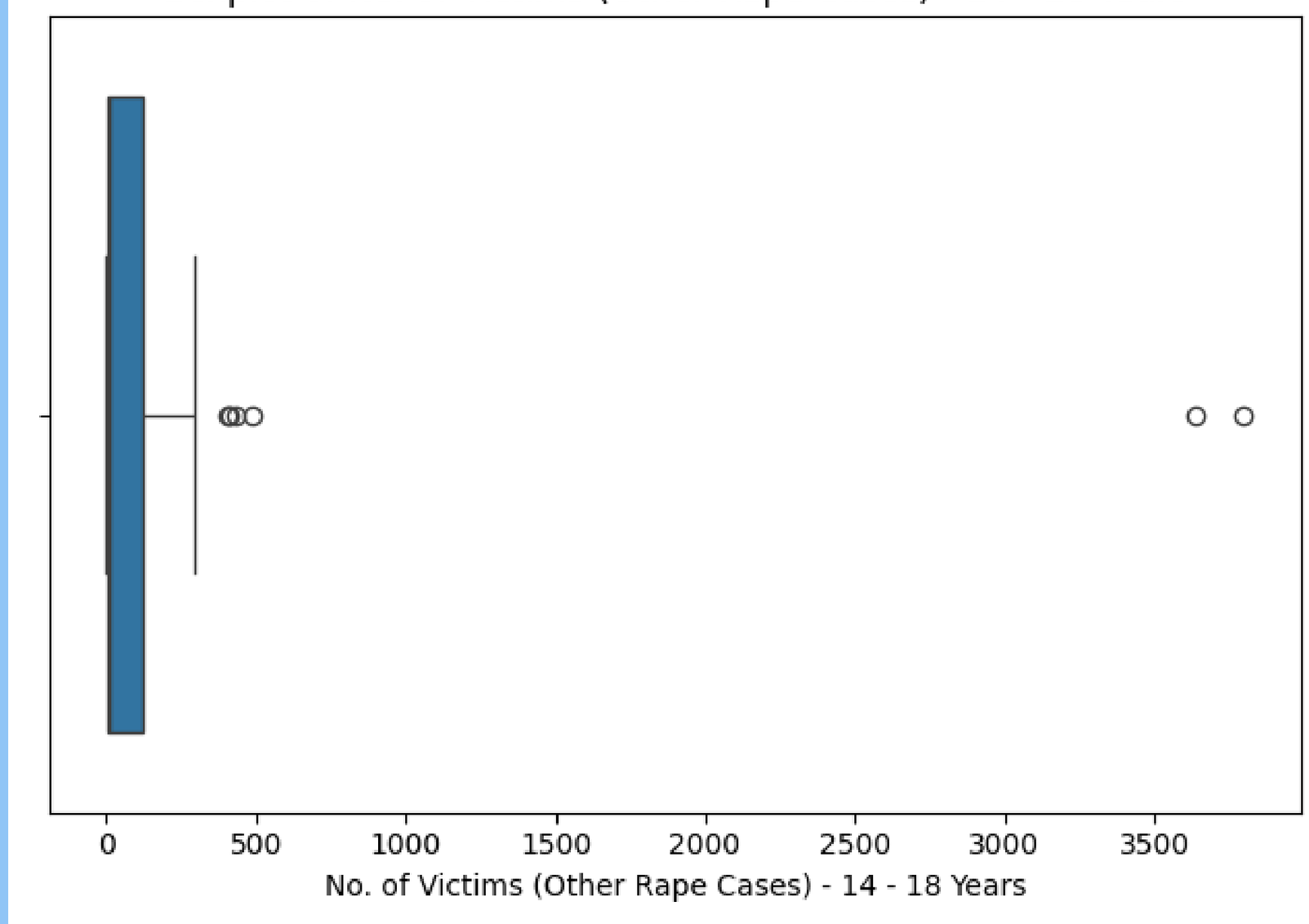

Boxplot of No. of Victims (Incest Rape Cases) - 30 - 50 Years
0
10
20
30
40
50
60
70
No. of Victims (Incest Rape Cases) - 30 - 50 Years

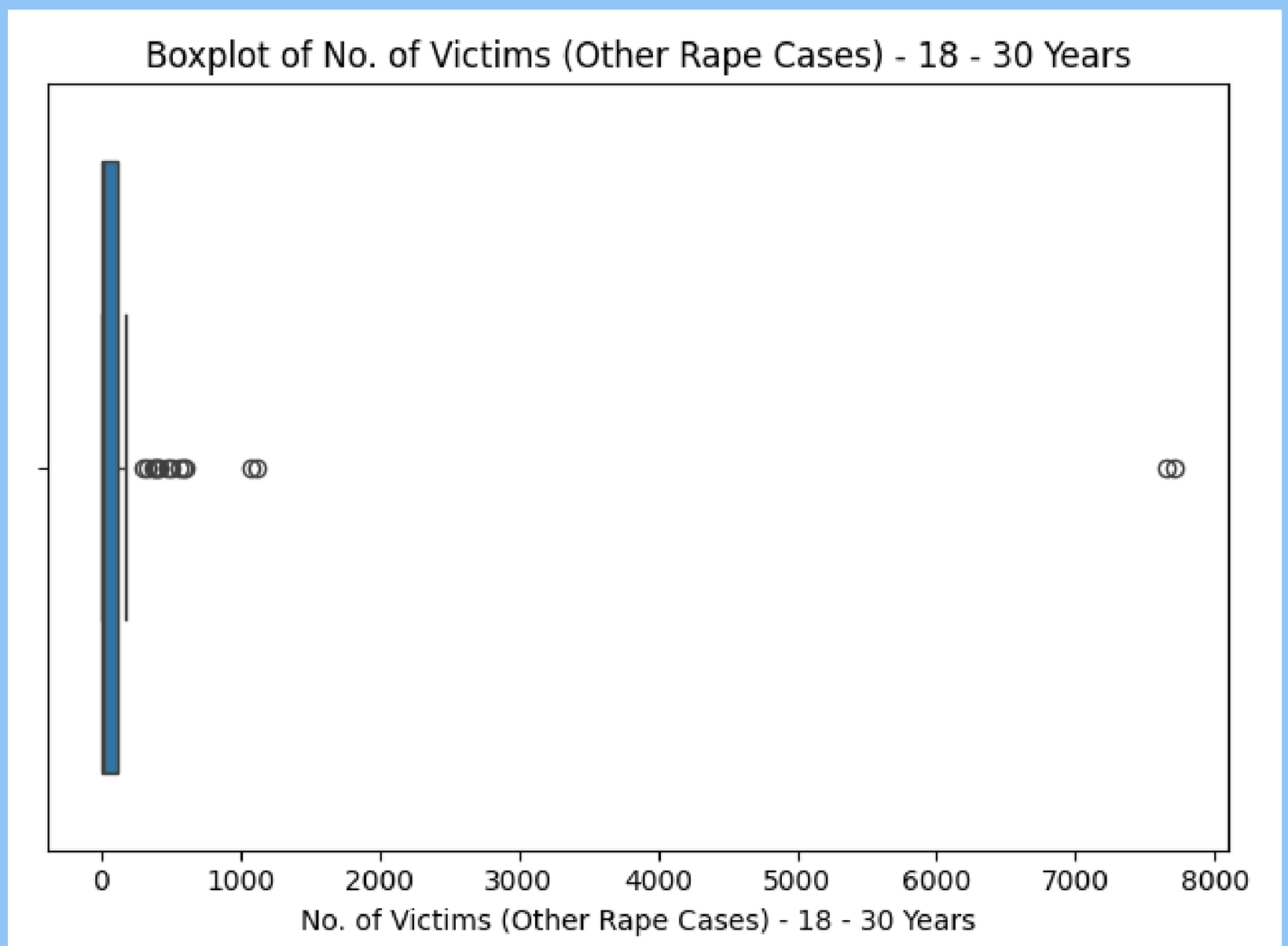

Boxplot of No. of Victims (Incest Rape Cases) - Total Victims
0
100
200
300
400
No. of Victims (Incest Rape Cases) - Total Victims

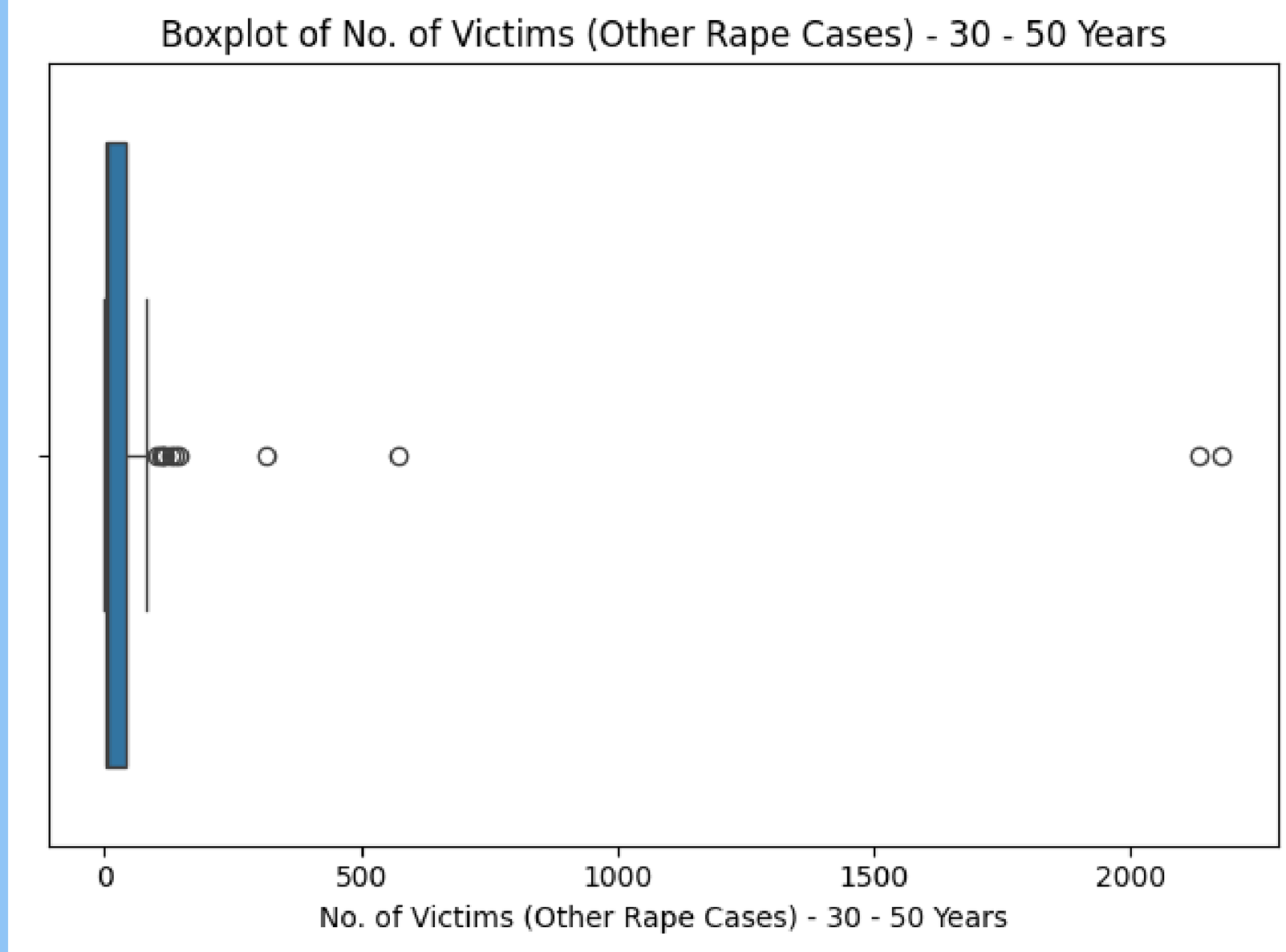

Boxplot of No. of Victims (Incest Rape Cases) above 50 Years
0.0
0.2
0.4
0.6
0.8
1.0
No. of Victims (Incest Rape Cases) above 50 Years

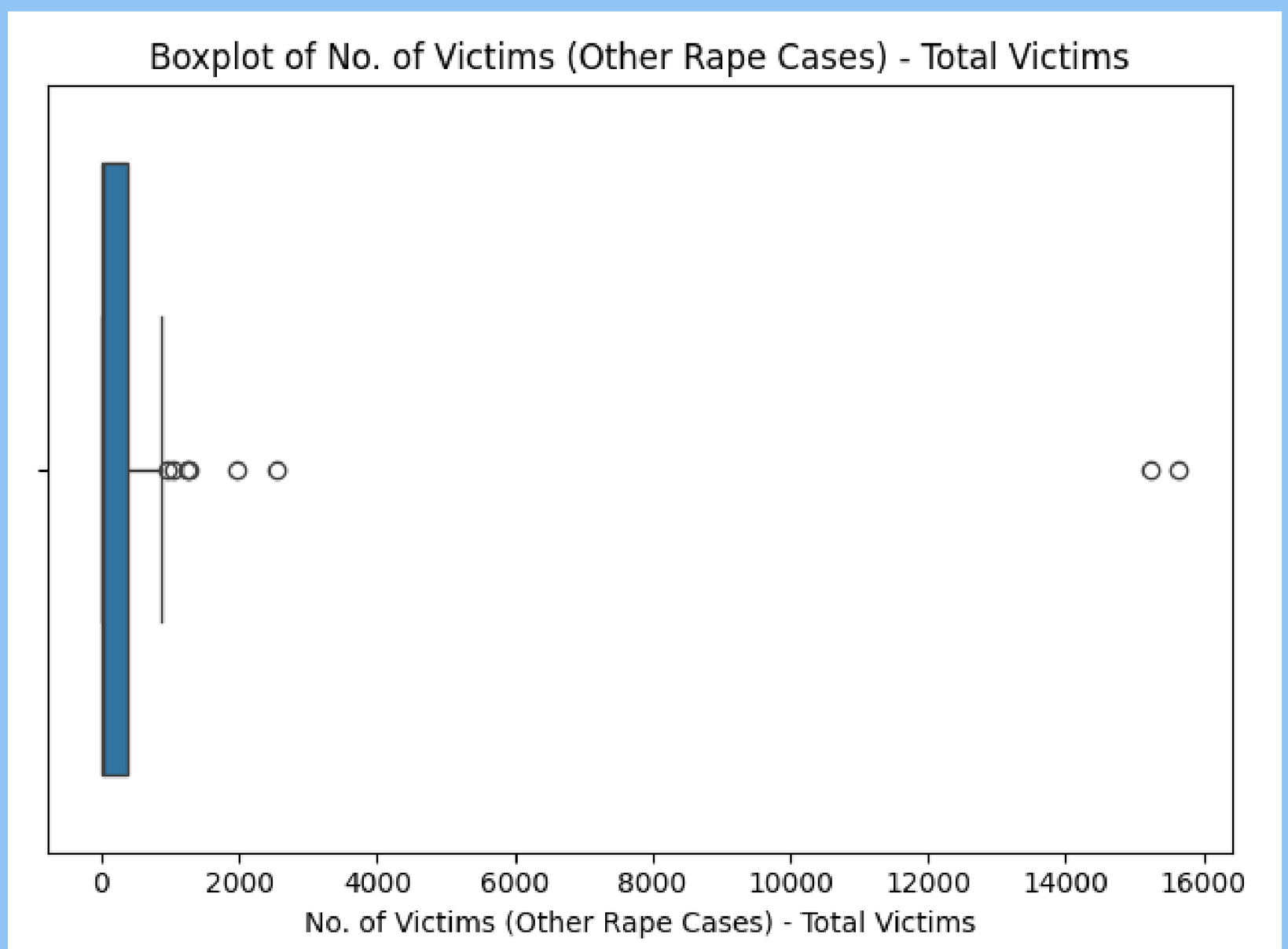

Boxplot of No. of Victims (Incest Rape Cases) upto 10 Years
0
5
10
15
20
25
No. of Victims (Incest Rape Cases) upto 10 Years

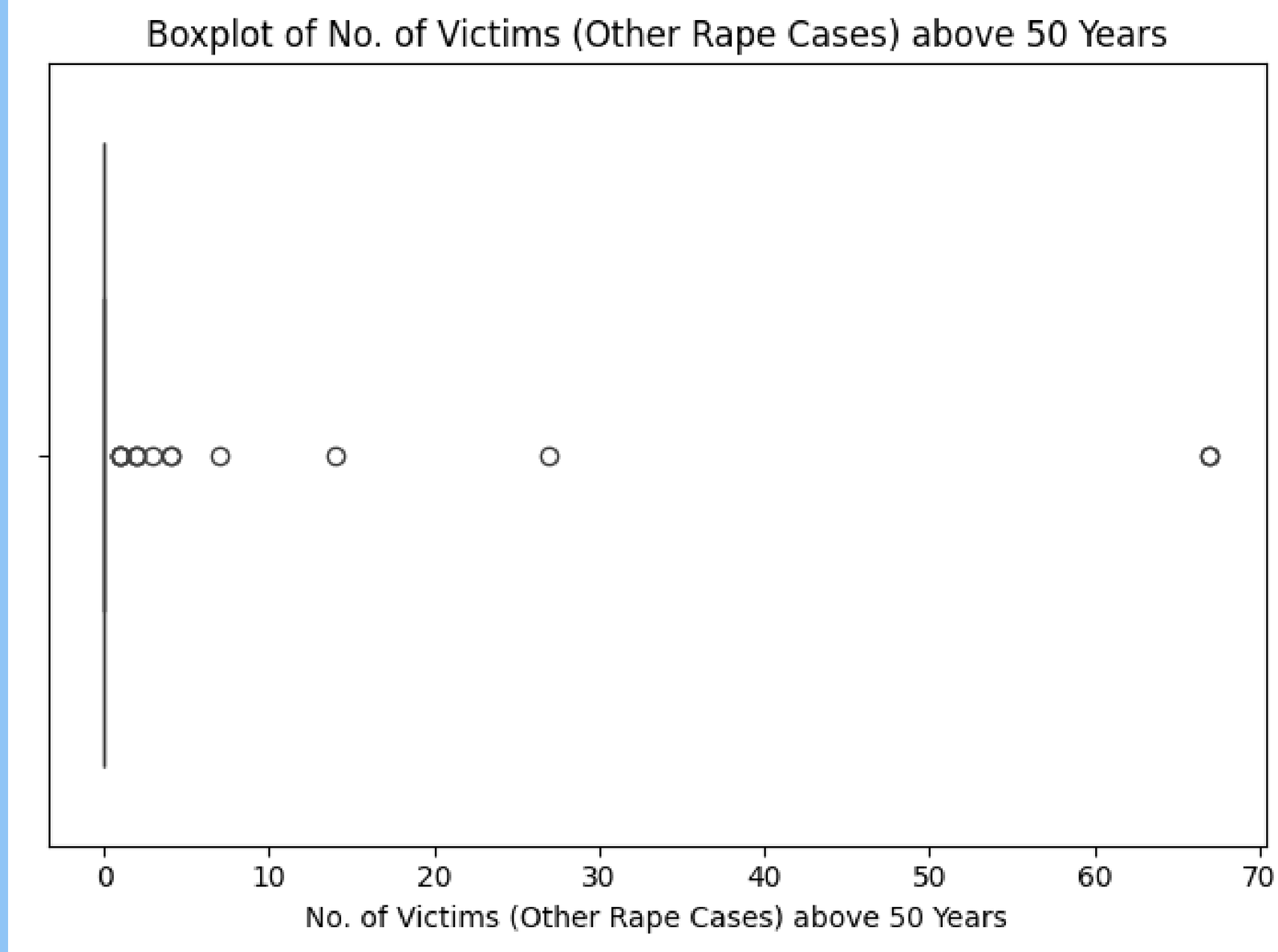

Boxplot of No. of Victims (Other Rape Cases) - 10 - 14 Years
0
200
400
600
800
1000
1200
1400
No. of Victims (Other Rape Cases) - 10 - 14 Years

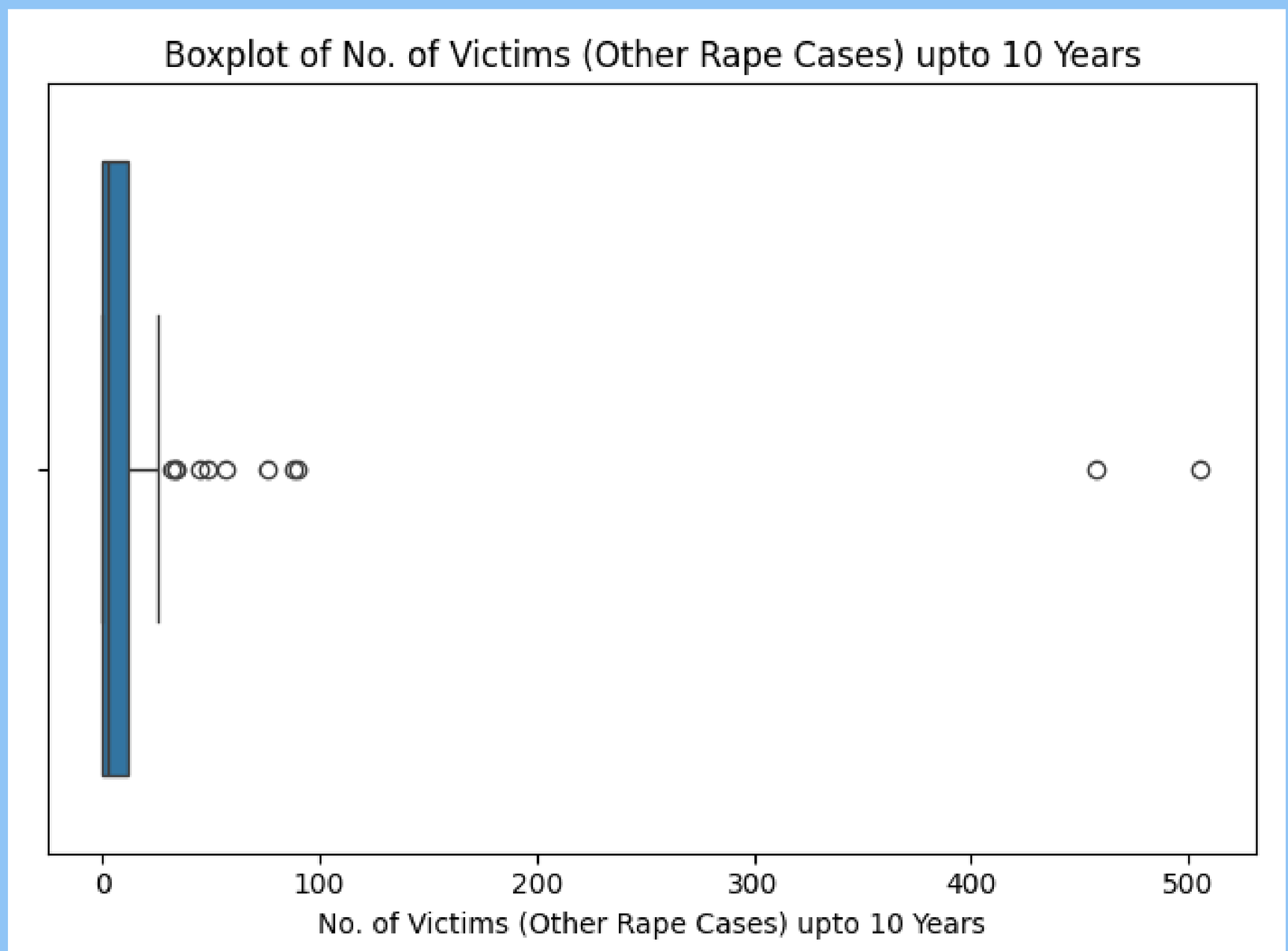

Boxplot of No. of Victims (Other Rape Cases) - 14 - 18 Years
0
500
1000
1500
2000
2500
3000
3500
No. of Victims (Other Rape Cases) - 14 - 18 Years

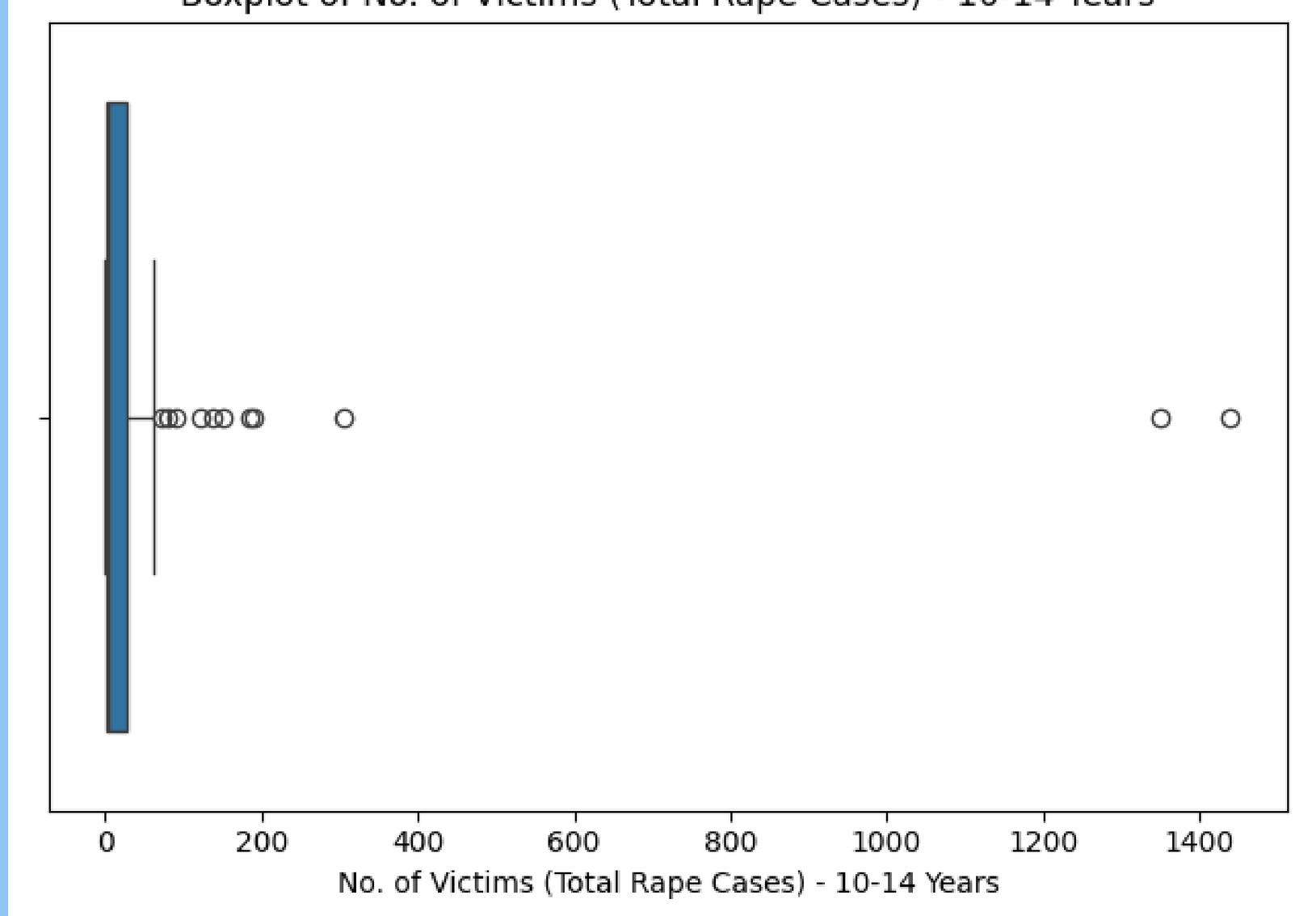

Boxplot of No. of Victims (Other Rape Cases) - 18 - 30 Years
0
1000
2000
3000
4000
5000
6000
7000
8000
No. of Victims (Other Rape Cases) - 18 - 30 Years

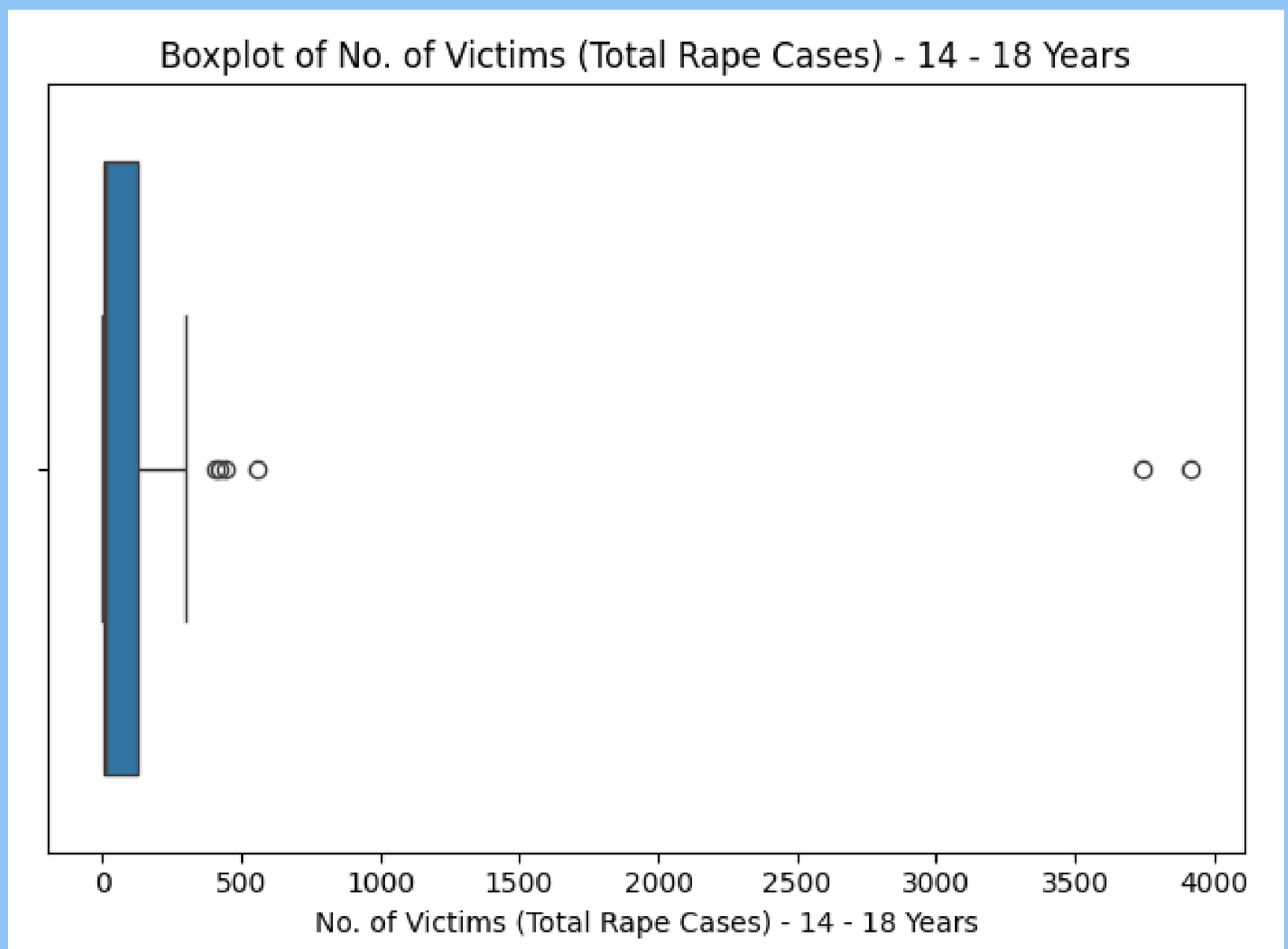

Boxplot of No. of Victims (Other Rape Cases) - 30 - 50 Years
0
500
1000
1500
2000
No. of Victims (Other Rape Cases) - 30 - 50 Years

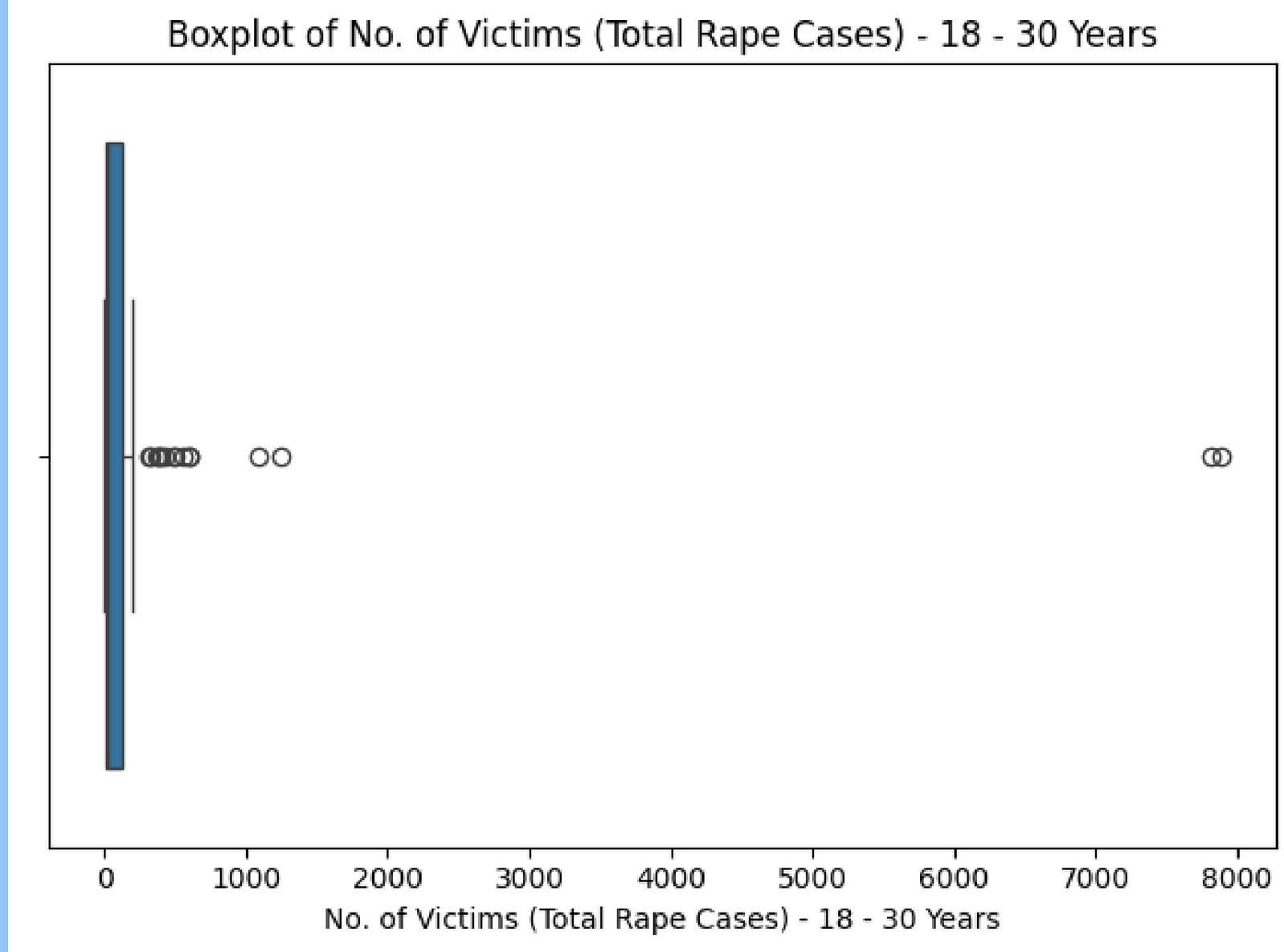

Boxplot of No. of Victims (Other Rape Cases) - Total Victims
0
2000
4000
6000
8000
10000
12000
14000
16000
No. of Victims (Other Rape Cases) - Total Victims


Boxplot of No. of Victims (Other Rape Cases) above 50 Years
0
10
20
30
40
50
60
70
No. of Victims (Other Rape Cases) above 50 Years

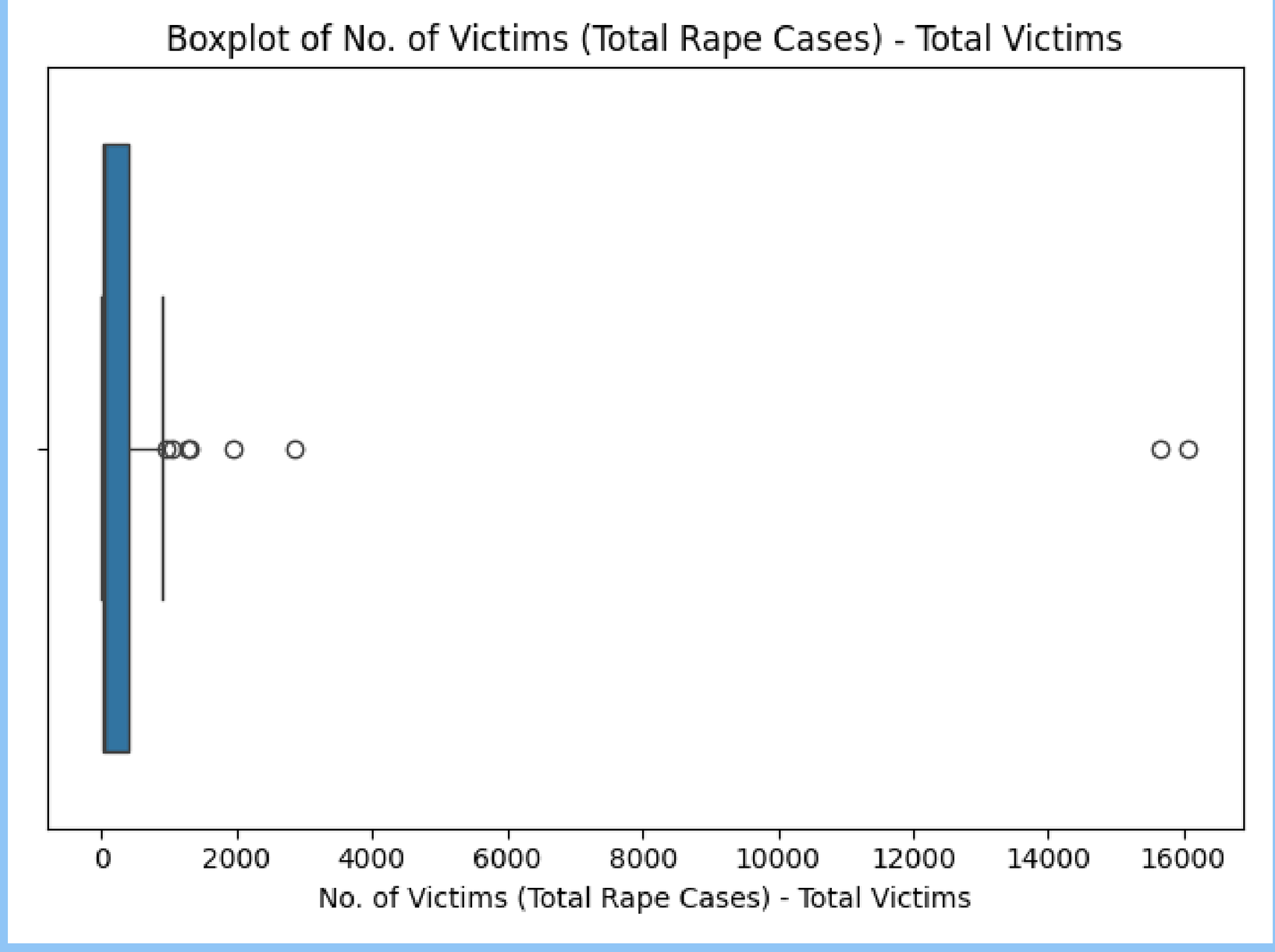

Boxplot of No. of Victims (Other Rape Cases) upto 10 Years
0
100
200
300
400
500
No. of Victims (Other Rape Cases) upto 10 Years

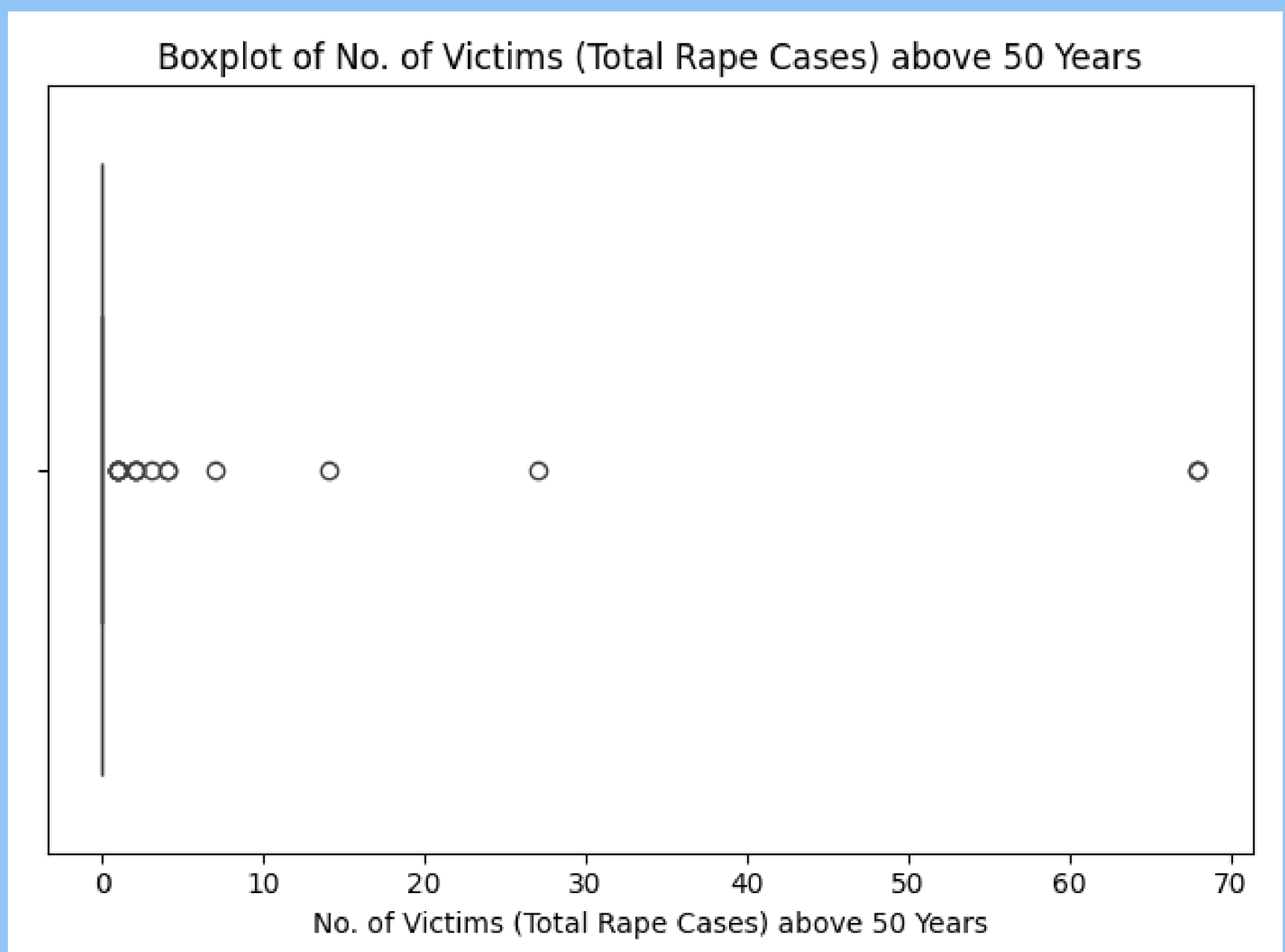

Boxplot of No. of Victims (Total Rape Cases) - 10-14 Years
0
200
400
600
800
1000
1200
1400
No. of Victims (Total Rape Cases) - 10-14 Years

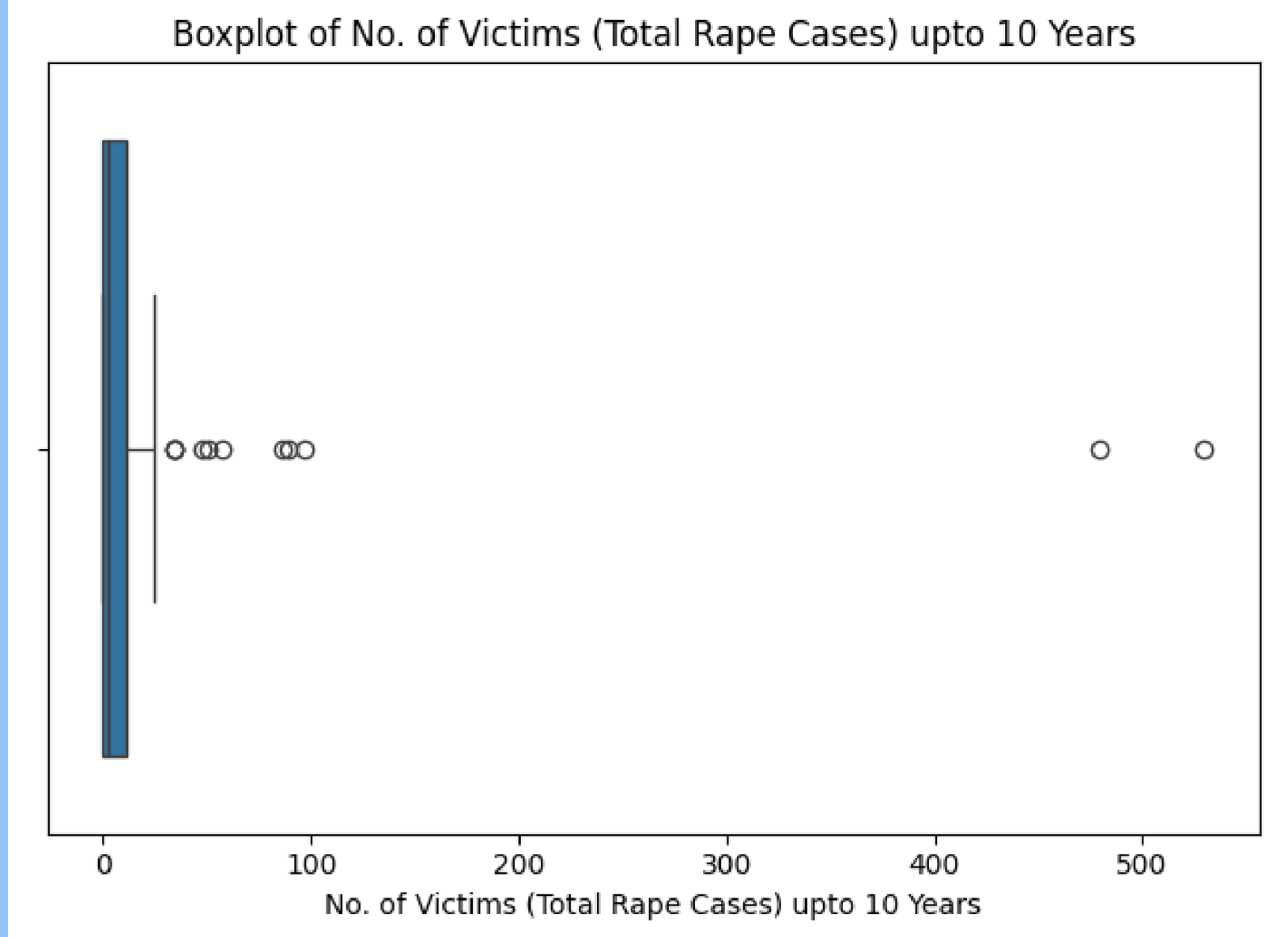

Boxplot of No. of Victims (Total Rape Cases) - 14 - 18 Years
0
500
1000
1500
2000
2500
3000
3500
4000
No. of Victims (Total Rape Cases) - 14 - 18 Years


Boxplot of No. of Victims (Total Rape Cases) - 18 - 30 Years
0
1000
2000
3000
4000
5000
6000
7000
8000
No. of Victims (Total Rape Cases) - 18 - 30 Years

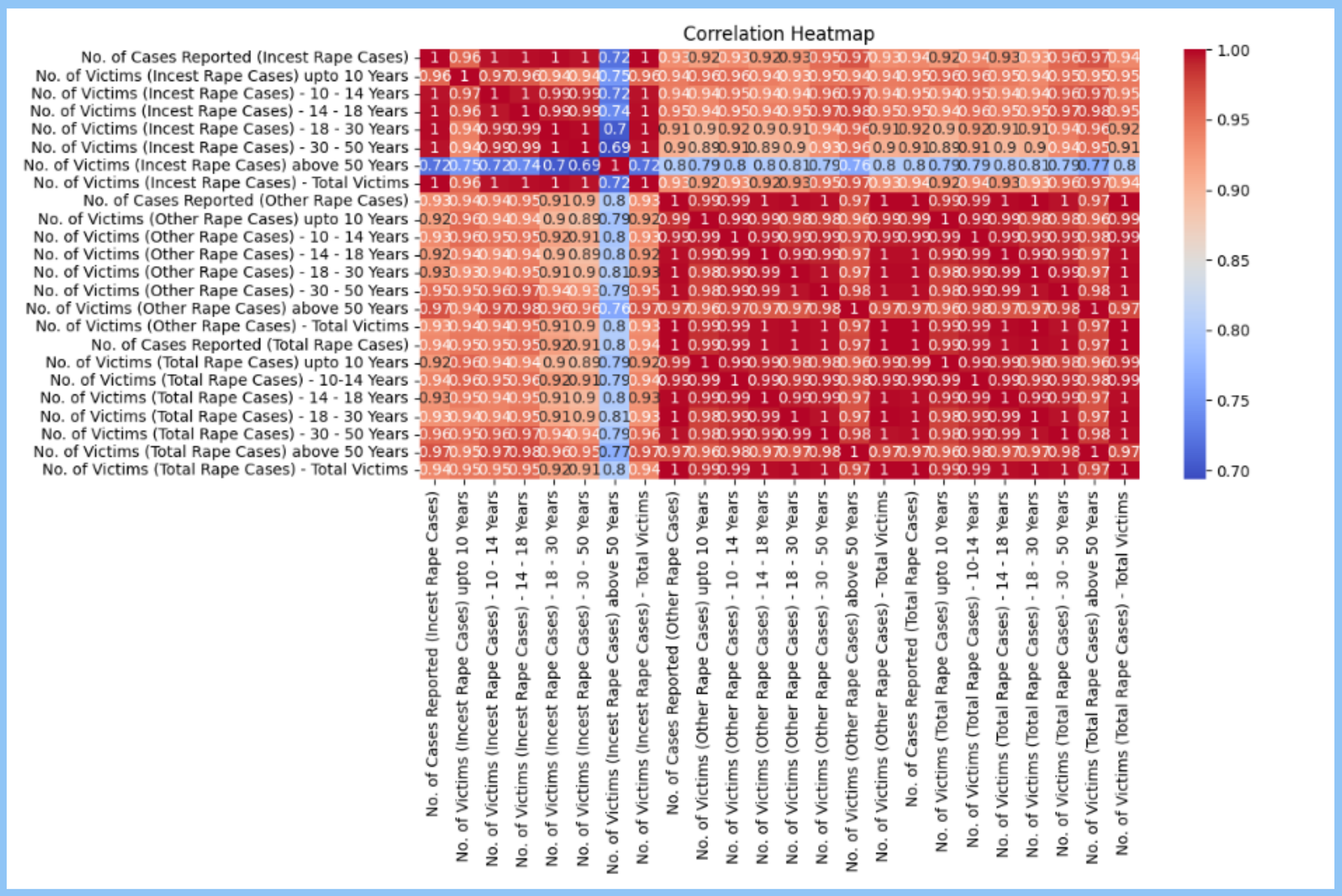
Boxplot of No. of Victims (Total Rape Cases) - 30 - 50 Years
0
500
1000
1500
2000
No. of Victims (Total Rape Cases) - 30 - 50 Years


Boxplot of No. of Victims (Total Rape Cases) - Total Victims
0
2000
4000
6000
8000
10000
12000
14000
16000
No. of Victims (Total Rape Cases) - Total Victims


Boxplot of No. of Victims (Total Rape Cases) above 50 Years
0
10
20
30
40
50
60
70
No. of Victims (Total Rape Cases) above 50 Years


Boxplot of No. of Victims (Total Rape Cases) upto 10 Years
0
100
200
300
400
500
No. of Victims (Total Rape Cases) upto 10 Years

Bivariate Analysis (Heatmap - Correlation Matrix)



4. To prepare the dataset for further analysis or modeling

Ready to proceed Data preprocessing and Data modeling.