In [1]:
```
#IMPORT THE NECCESSARY LIBRARY
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from tqdm.auto import tqdm
```

In [2]:
```
df=pd.read_csv('Table43.3-State.csv')
df.head(35)
```

Out[2]:

	States/UTs	Number of constituencies - Total constituencies	Number of constituencies - Elected Women	Number of electors - Total	Number of electors - Women Electors	Number of votes polled - Total	Number of votes polled - Women	Percentage of Elected women over total seats	Percentage of Women electors over total electors	Percentage of Women's votes over total votes polled	Percentage of Women voters over women electors	Unnamed: 11	Unnamed: 12	Unnamed: 13	Unnamed: 14	Unnamed: 15	Unnamed: 16	Unnamed: 17	Unnamed: 18
0	Andhra Pradesh	42.0	5.0	57892259.0	29207418.0	42048269.0	20845852.0	11.9	50.5	49.6	71.4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Arunachal Pradesh	2.0	NaN	734541.0	364877.0	500642.0	245284.0	NaN	49.7	49.0	67.2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	Assam	14.0	2.0	17470329.0	8439234.0	12147016.0	5632823.0	14.3	48.3	46.4	66.7	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Bihar	40.0	4.0	54505246.0	25284439.0	24235476.0	10775241.0	10.0	46.4	44.5	42.6	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	Chattisgarh	11.0	2.0	15476577.0	7626789.0	8555661.0	3983764.0	18.2	49.3	46.6	52.2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	Goa	2.0	NaN	1020794.0	508319.0	564255.0	273482.0	NaN	49.8	48.5	53.8	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
6	Gujarat	26.0	4.0	36484281.0	17585894.0	17476088.0	7624781.0	15.4	48.2	43.6	43.4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7	Haryana	10.0	2.0	12087710.0	5496756.0	8157676.0	3619042.0	20.0	45.5	44.4	65.8	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8	Himachal Pradesh	4.0	NaN	4606674.0	2257953.0	2690850.0	1334297.0	NaN	49.0	49.6	59.1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9	Jammu & Kashmir	6.0	NaN	6572896.0	3151188.0	2607880.0	1065887.0	NaN	47.9	40.9	33.8	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
10	Jharkhand	14.0	NaN	17934095.0	8449928.0	9142173.0	4033059.0	NaN	47.1	44.1	47.7	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11	Karnataka	28.0	1.0	41790939.0	20474457.0	24575813.0	11592518.0	3.6	49.0	47.2	56.6	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
12	Kerala	20.0	NaN	21859536.0	11330955.0	16036873.0	8220848.0	NaN	51.8	51.3	72.6	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
13	Madhya Pradesh	29.0	6.0	38085179.0	17902080.0	19486218.0	7852392.0	20.7	47.0	40.3	43.9	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
14	Maharashtra	48.0	3.0	72954058.0	34793896.0	36995037.0	16487190.0	6.3	47.7	44.6	47.4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
15	Manipur	2.0	NaN	1736251.0	899752.0	1339399.0	690644.0	NaN	51.8	51.6	76.8	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
16	Meghalaya	2.0	1.0	1277739.0	648303.0	822566.0	415923.0	50.0	50.7	50.6	64.2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
17	Mizoram	1.0	NaN	629374.0	318227.0	325991.0	154086.0	NaN	50.6	47.3	48.4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
18	Nagaland	1.0	NaN	1321878.0	645156.0	1189601.0	575635.0	NaN	48.8	48.4	89.2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
19	Odisha	21.0	NaN	27194864.0	13119010.0	17767143.0	8452619.0	NaN	48.2	47.6	64.4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
20	Punjab	13.0	4.0	16958380.0	8129384.0	11832306.0	5645170.0	30.8	47.9	47.7	69.4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
21	Rajasthan	25.0	3.0	37060011.0	17520650.0	17935280.0	7844124.0	12.0	47.3	43.7	44.8	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
22	Sikkim	1.0	NaN	300584.0	143222.0	251776.0	118668.0	NaN	47.6	47.1	82.9	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
23	Tamil Nadu	39.0	1.0	41620460.0	20725138.0	30397137.0	14903459.0	2.6	49.8	49.0	71.9	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
24	Tripura	2.0	NaN	2082265.0	1016559.0	1758501.0	838874.0	NaN	48.8	47.7	82.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25	Uttar Pradesh	80.0	13.0	116006374.0	52611874.0	55430198.0	23271819.0	16.3	45.4	42.0	44.2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
26	Uttarakhand	5.0	NaN	5887724.0	2850182.0	3140699.0	1441326.0	NaN	48.4	45.9	50.6	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
27	West Bengal	42.0	7.0	52493168.0	24925428.0	42731483.0	20003460.0	16.7	47.5	46.8	80.3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
28	Andaman & Nicobar Islands	1.0	NaN	265108.0	123959.0	170103.0	78359.0	NaN	46.8	46.1	63.2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
29	Chandigarh	1.0	NaN	524444.0	232778.0	343557.0	151176.0	NaN	44.4	44.0	64.9	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
30	Dadra & Nagar Haveli	1.0	NaN	150704.0	69643.0	110363.0	52617.0	NaN	46.2	47.7	75.6	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
31	Daman & Diu	1.0	NaN	95382.0	47025.0	68024.0	35522.0	NaN	49.3	52.2	75.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
32	Delhi	7.0	1.0	11096854.0	4907979.0	5754256.0	2433896.0	14.3	44.2	42.3	49.6	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
33	Lakshadweep	1.0	NaN	45983.0	22481.0	39498.0	19777.0	NaN	48.9	50.1	88.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
34	Puduchery	1.0	NaN	762440.0	395367.0	608503.0	315383.0	NaN	51.9	51.8	79.8	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [3]:
```
# Remove all unnamed columns
unnamed_columns = df.filter(regex='Unnamed').columns
df.drop(unnamed_columns, axis=1, inplace=True)

# Drop rows with any NaN values
# Since we want to keep the table as it is, we will not impute missing values
df.dropna(inplace=True)

# Display the cleaned dataframe
print(df.head())
```

```
       States/UTs  Number of constituencies - Total constituencies  \
0  Andhra Pradesh                                             42.0
2           Assam                                             14.0
3           Bihar                                             40.0
4     Chattisgarh                                             11.0
6         Gujarat                                             26.0

   Number of constituencies - Elected Women  Number of electors - Total  \
0                                       5.0                   57892259.0
2                                       2.0                   17470329.0
3                                       4.0                   54505246.0
4                                       2.0                   15476577.0
6                                       4.0                   36484281.0

   Number of electors - Women Electors  Number of votes polled - Total  \
0                           29207418.0                      42048269.0
2                            8439234.0                      12147016.0
3                           25284439.0                      24235476.0
4                            7626789.0                       8555661.0
6                           17585894.0                      17476088.0

   Number of votes polled - Women  \
0                      20845852.0
2                       5632823.0
3                      10775241.0
4                       3983764.0
6                       7624781.0

   Percentage of Elected women over total seats  \
0                                          11.9
2                                          14.3
3                                          10.0
4                                          18.2
6                                          15.4

   Percentage of Women electors over total electors  \
0                                              50.5
2                                              48.3
3                                              46.4
4                                              49.3
6                                              48.2

   Percentage of Women's votes over total votes polled  \
0                                               49.6
2                                               46.4
3                                               44.5
4                                               46.6
6                                               43.6

   Percentage of Women voters over women electors
0                                            71.4
2                                            66.7
3                                            42.6
4                                            52.2
6                                            43.4
```

This code snippet in Python using Pandas performs two operations on a DataFrame (`df`):

1. **`unnamed_columns = df.filter(regex='Unnamed').columns`** :
   - `df.filter(regex='Unnamed')` : Filters the DataFrame `df` to select columns where the column names match the regular expression pattern `'Unnamed'` .
   - `.columns` : Retrieves the column labels from the filtered DataFrame.
   - `unnamed_columns` : Stores the column labels of columns that match the `'Unnamed'` pattern.
2. **`df.drop(unnamed_columns, axis=1, inplace=True)`** :
   - `df.drop()` : Drops columns from the DataFrame.
   - `unnamed_columns` : Specifies the columns to be dropped, which are the columns with names matching the `'Unnamed'` pattern.
   - `axis=1` : Specifies that the operation should be performed along columns (1 indicates columns, while 0 would indicate rows).
   - `inplace=True` : Modifies the DataFrame `df` in place without creating a new DataFrame.

Overall, this code snippet identifies columns in the DataFrame `df` whose names contain the string 'Unnamed' using a regular expression. It then drops these columns from the DataFrame, effectively removing any columns with names containing 'Unnamed'. This is often used when reading data from files like CSV, where unnamed or extra columns without specific names might be generated.

In [4]:
```
# Calculate summary statistics for the numerical columns
df.describe()
```

Out[4]:

	Number of constituencies - Total constituencies	Number of constituencies - Elected Women	Number of electors - Total	Number of electors - Women Electors	Number of votes polled - Total	Number of votes polled - Women	Percentage of Elected women over total seats	Percentage of Women electors over total electors	Percentage of Women's votes over total votes polled	Percentage of Women voters over women electors
count	16.000000	16.000000	1.600000e+01	1.600000e+01	1.600000e+01	1.600000e+01	16.000000	16.000000	16.000000	16.000000
mean	28.500000	3.687500	3.895372e+07	1.851748e+07	2.241128e+07	1.018322e+07	16.443750	47.793750	45.581250	57.150000
std	19.869575	3.092329	2.865466e+07	1.327046e+07	1.545344e+07	7.030984e+06	11.307105	1.839735	2.896025	12.637932
min	2.000000	1.000000	1.277739e+06	6.483030e+05	8.225660e+05	4.159230e+05	2.600000	44.200000	40.300000	42.600000
25%	12.500000	1.750000	1.658793e+07	8.003735e+06	1.101314e+07	5.220558e+06	11.425000	46.850000	43.675000	44.650000
50%	27.000000	3.000000	3.757260e+07	1.774399e+07	1.871075e+07	7.848258e+06	14.850000	47.800000	45.500000	54.400000
75%	40.500000	4.250000	5.299619e+07	2.501518e+07	3.204661e+07	1.529939e+07	18.650000	49.075000	47.325000	67.375000
max	80.000000	13.000000	1.160064e+08	5.261187e+07	5.543020e+07	2.327182e+07	50.000000	50.700000	50.600000	80.300000

INFERENCE:

The summary statistics for the numerical columns of the dataset provide the following insights: The average number of total constituencies per state/UT is approximately 28.5, with a standard deviation of about 19.8, indicating a wide variance in the number of constituencies across different states/UTs.

On average, there are about 3.7 elected women per state/UT among those with participated women, with a standard deviation of approximately 3.1, suggesting that the number of elected women also varies significantly across different states/UTs

The mean percentage of elected women over total seats is 16.44%, with some states/UTs having a higher representation of women and others having less.

The average percentage of women electors over total electors is about 48.33%, suggesting that women constitute nearly half of the electorate on average

The average percentage of women's votes over total votes polled is approximately 46.86%, indicating that women's participation in voting is slightly less than their representation in the electorate

These statistics highlight the variability and distribution of electoral data across different states/UTs, as well as the representation and participation of women in the electoral process.

In [5]:
```
# Exclude the 'States/UTs' column from the calculations
numeric_columns = df.select_dtypes(include=['number'])

# Calculate median, mode, and range for the numerical columns
median_values = numeric_columns.median()
mode_values = numeric_columns.mode().iloc[0]
range_values = numeric_columns.max() - numeric_columns.min()

median_values, mode_values, range_values
```

Out[5]:
```
(Number of constituencies - Total constituencies             27.00
 Number of constituencies - Elected Women                     3.00
 Number of electors - Total                            37572595.00
 Number of electors - Women Electors                   17743987.00
 Number of votes polled - Total                        18710749.00
 Number of votes polled - Women                         7848258.00
 Percentage of Elected women over total seats                 14.85
 Percentage of Women electors over total electors             47.80
 Percentage of Women's votes over total votes polled          45.50
 Percentage of Women voters over women electors               54.40
 dtype: float64,
 Number of constituencies - Total constituencies             42.0
 Number of constituencies - Elected Women                     1.0
 Number of electors - Total                             1277739.0
 Number of electors - Women Electors                     648303.0
 Number of votes polled - Total                          822566.0
 Number of votes polled - Women                          415923.0
 Percentage of Elected women over total seats                14.3
 Percentage of Women electors over total electors            44.2
 Percentage of Women's votes over total votes polled         40.3
 Percentage of Women voters over women electors              42.6
 Name: 0, dtype: float64,
 Number of constituencies - Total constituencies             78.0
 Number of constituencies - Elected Women                    12.0
 Number of electors - Total                            114728635.0
 Number of electors - Women Electors                    51963571.0
 Number of votes polled - Total                         54607632.0
 Number of votes polled - Women                         22855896.0
 Percentage of Elected women over total seats                47.4
 Percentage of Women electors over total electors             6.5
 Percentage of Women's votes over total votes polled         10.3
 Percentage of Women voters over women electors              37.7
 dtype: float64)
```

INFERENCE:

Number of constituencies - Total constituencies - negative skewed. (median<mode)

Number of constituencies - Elected Women - positive skewed.(median>mode)

Number of electors - Total - positive skewed.(median>mode)

Number of electors - Women Electors- positive skewed.(median>mode)

similarly,with all the other columns

In [6]:
```
# Calculate skewness and kurtosis for the numerical columns
skewness_values = numeric_columns.skew()
kurtosis_values = numeric_columns.kurtosis()
skewness_values, kurtosis_values
```

Out[6]:
```
(Number of constituencies - Total constituencies        1.015305
 Number of constituencies - Elected Women                1.973559
 Number of electors - Total                              1.236522
 Number of electors - Women Electors                     1.039499
 Number of votes polled - Total                          0.678188
 Number of votes polled - Women                          0.572581
 Percentage of Elected women over total seats            1.822825
 Percentage of Women electors over total electors       -0.217053
 Percentage of Women's votes over total votes polled     0.004986
 Percentage of Women voters over women electors          0.327485
 dtype: float64,
 Number of constituencies - Total constituencies        1.593535
 Number of constituencies - Elected Women               4.905788
 Number of electors - Total                             2.305326
 Number of electors - Women Electors                    1.557817
 Number of votes polled - Total                        -0.294706
 Number of votes polled - Women                        -0.828399
 Percentage of Elected women over total seats           4.751678
 Percentage of Women electors over total electors      -0.403061
 Percentage of Women's votes over total votes polled   -0.658998
 Percentage of Women voters over women electors        -1.403064
 dtype: float64)
```

INFERENCE

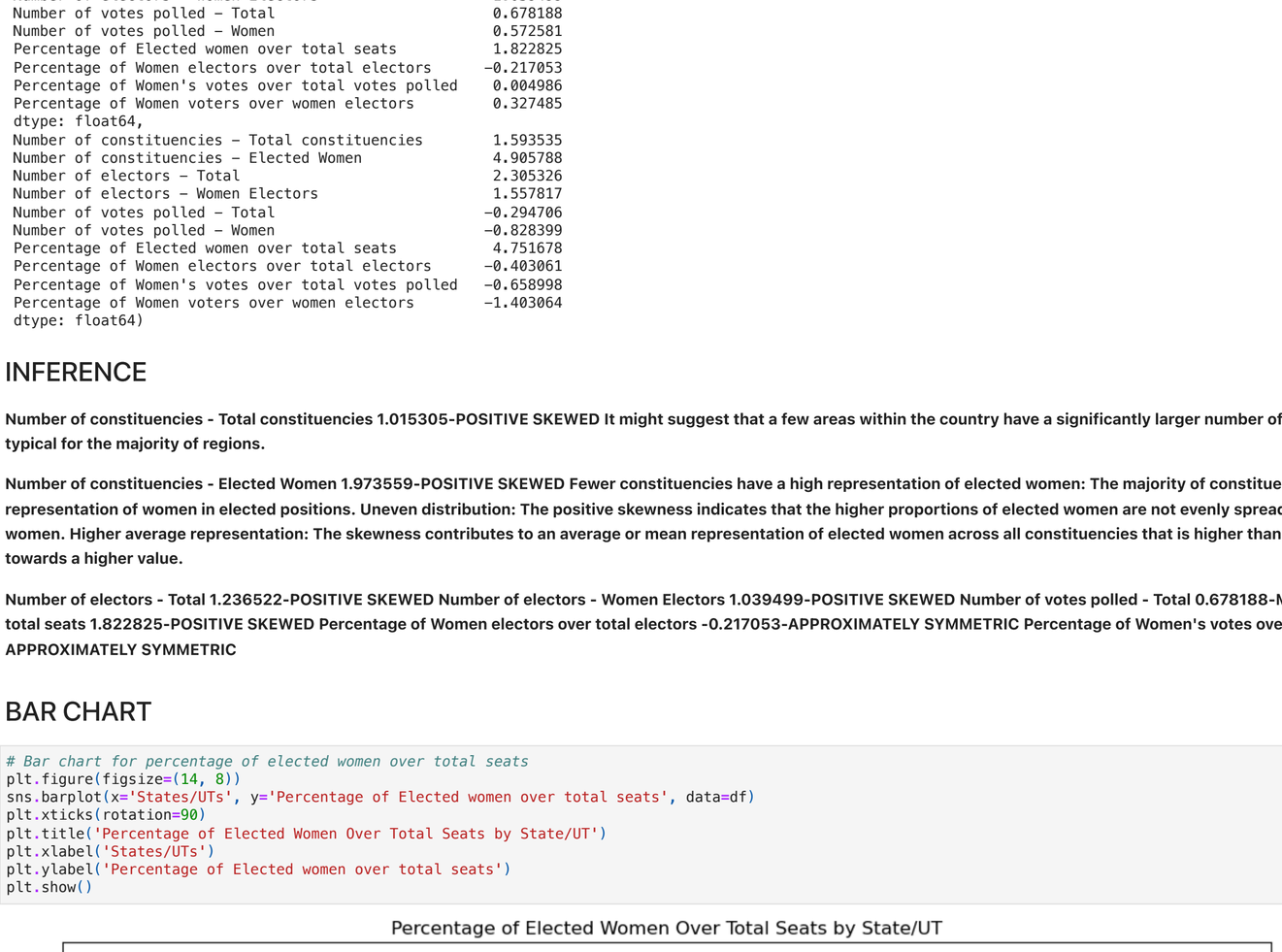
Number of constituencies - Total constituencies 1.015305-POSITIVE SKEWED It might suggest that a few areas within the country have a significantly larger number of constituencies compared to most other areas, causing the average or mean number of constituencies to be higher than what is typical for the majority of regions.

Number of constituencies - Elected Women 1.973559-POSITIVE SKEWED Fewer constituencies have a high representation of elected women: The majority of constituencies likely have a lower proportion of elected women compared to a few constituencies that boast a significantly higher representation of women in elected positions. Uneven distribution: The positive skewness indicates that the higher proportions of elected women are not evenly spread across all constituencies. Instead, there are relatively fewer constituencies with a notably higher representation of elected women. Higher average representation: The skewness contributes to an average or mean representation of elected women across all constituencies that is higher than the median representation. This is because the few constituencies with significantly more elected women skew the average towards a higher value.

Number of electors - Total 1.236522-POSITIVE SKEWED Number of electors - Women Electors 1.039499-POSITIVE SKEWED Number of votes polled - Total 0.678188-MODERATELY SKEWED Number of votes polled - Women 0.572581-MODERATELY SKEWED Percentage of Elected women over total seats 1.822825-POSITIVE SKEWED Percentage of Women electors over total electors -0.217053-APPROXIMATELY SYMMETRIC Percentage of Women's votes over total votes polled 0.004986-APPROXIMATELY SYMMETRIC Percentage of Women voters over women electors 0.327485-APPROXIMATELY SYMMETRIC

BAR CHART

In [7]:
```
# Bar chart for percentage of elected women over total seats
plt.figure(figsize=(14, 8))
sns.barplot(x='States/UTs', y='Percentage of Elected women over total seats', data=df)
plt.xticks(rotation=90)
plt.title('Percentage of Elected Women Over Total Seats by State/UT')
plt.xlabel('States/UTs')
plt.ylabel('Percentage of Elected women over total seats')
plt.show()
```



INFERENCE:

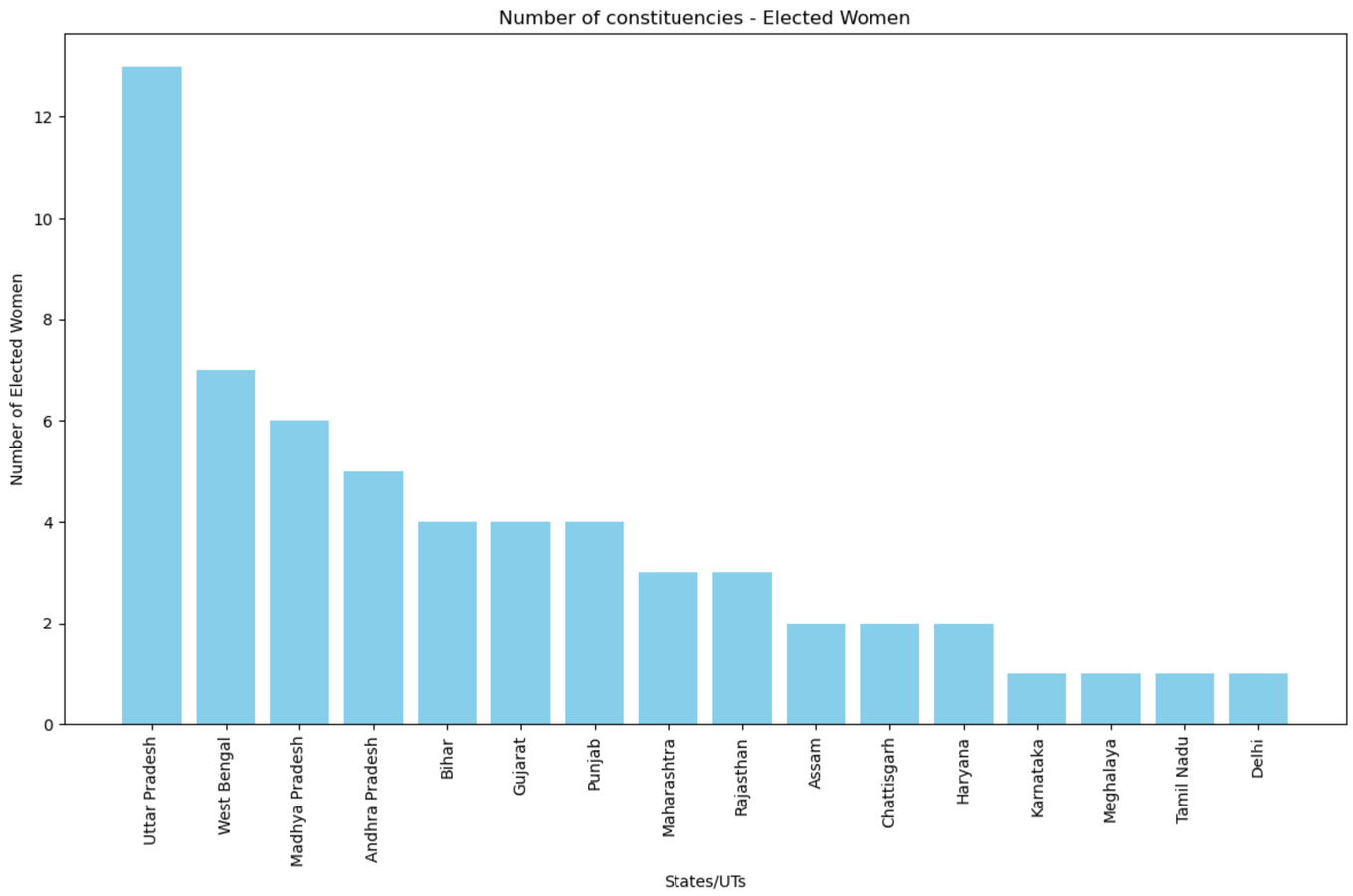
Percentage of Elected Women over Total Seats" focuses on the overall proportion of women in elected positions compared to the total number of available positions or seats in the entire system.

Percentage of Elected Women Over Total Seats by State/UT is highest in Meghalaya , followed by Punjab and Madhya Pradesh. It is least is Tamil Nadu, Karnataka and Maharashtra.Other states have moderate percentage of Elected Women Over Total Seats by State/UT.

In [8]:
```
# Select relevant columns
relevant_df = df[['States/UTs', 'Number of constituencies - Elected Women']].dropna()

# Sort the values by the number of elected women
relevant_df = relevant_df.sort_values(by='Number of constituencies - Elected Women', ascending=False)

# Plot
plt.figure(figsize=(12, 8))
plt.bar(relevant_df['States/UTs'], relevant_df['Number of constituencies - Elected Women'], color='skyblue')
plt.xlabel('States/UTs')
plt.ylabel('Number of Elected Women')
plt.xticks(rotation=90)
plt.title('Number of constituencies - Elected Women')
plt.tight_layout()
plt.savefig('elected_women_by_state.png')
plt.show()
```



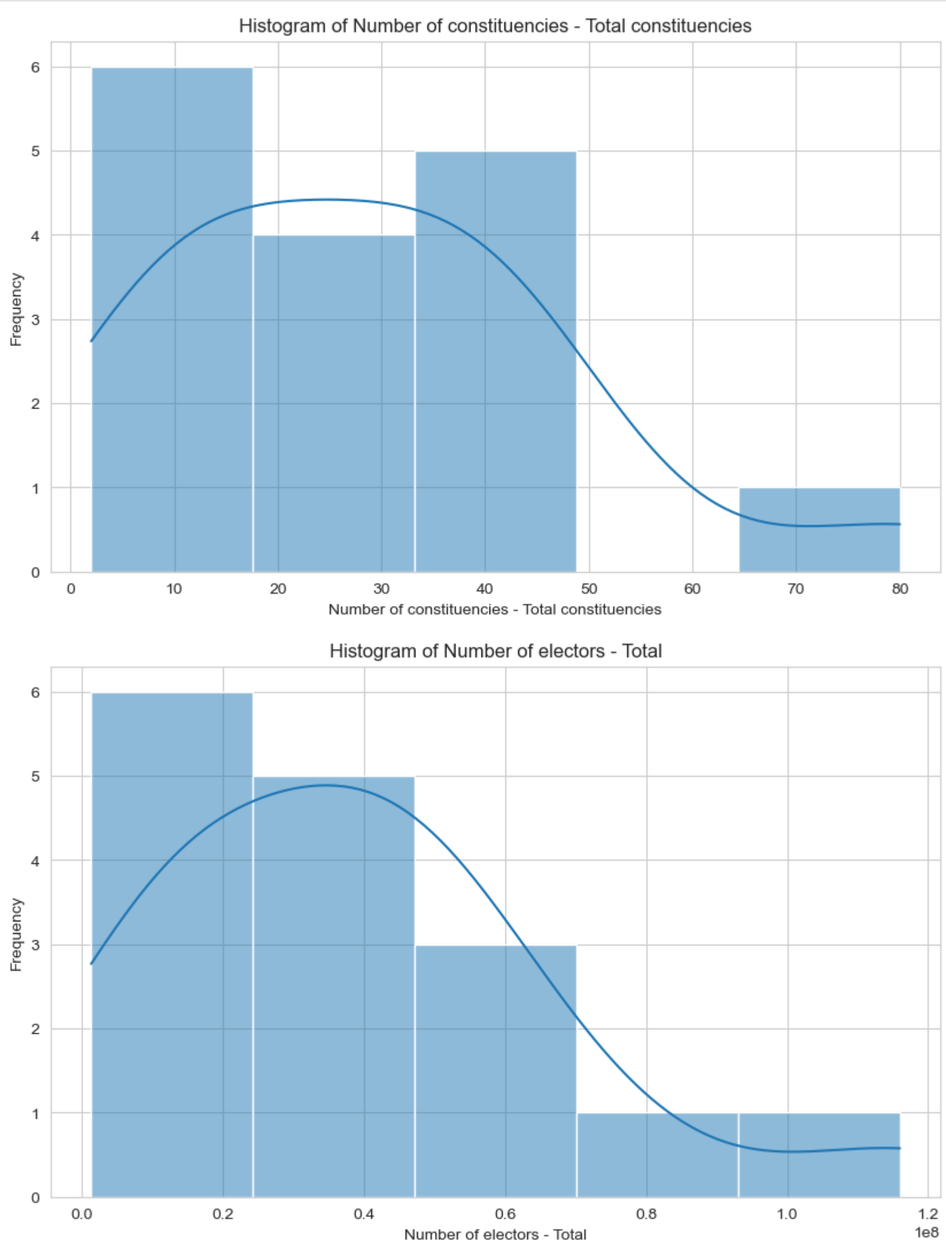
INFERENCE:Bar Chart of Elected Women by State/UT:

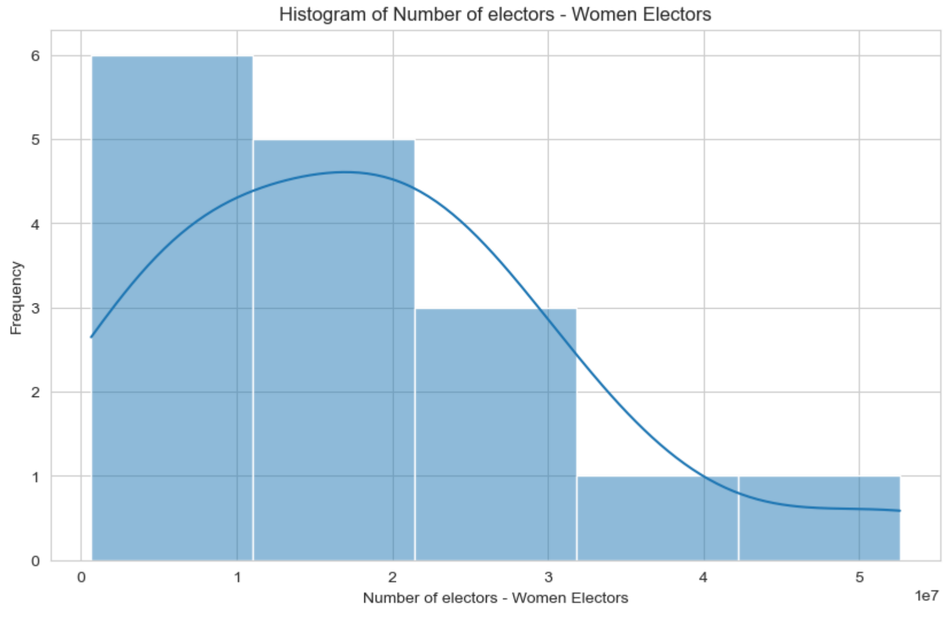
Number of Constituencies - Elected Women highlights the count or quantity of constituencies where women have been elected without necessarily considering the overall number of available seats.

1.Uttar Pradesh has the highest number of elected women, followed by West Bengal and Maharashtra. 2.States like Sikkim, Nagaland, and Mizoram have the lowest number of elected women.

In [9]:
```
# Set the aesthetic style of the plots
sns.set_style('whitegrid')

# Histograms for constituencies and electors
df_hist_columns = ['Number of constituencies - Total constituencies', 'Number of electors - Total', 'Number of electors - Women Electors']
for column in df_hist_columns:
    plt.figure(figsize=(10, 6))
    sns.histplot(df[column], kde=True)
    plt.title('Histogram of ' + column)
    plt.xlabel(column)
    plt.ylabel('Frequency')
    plt.show()
```





Conclusions from these histograms:

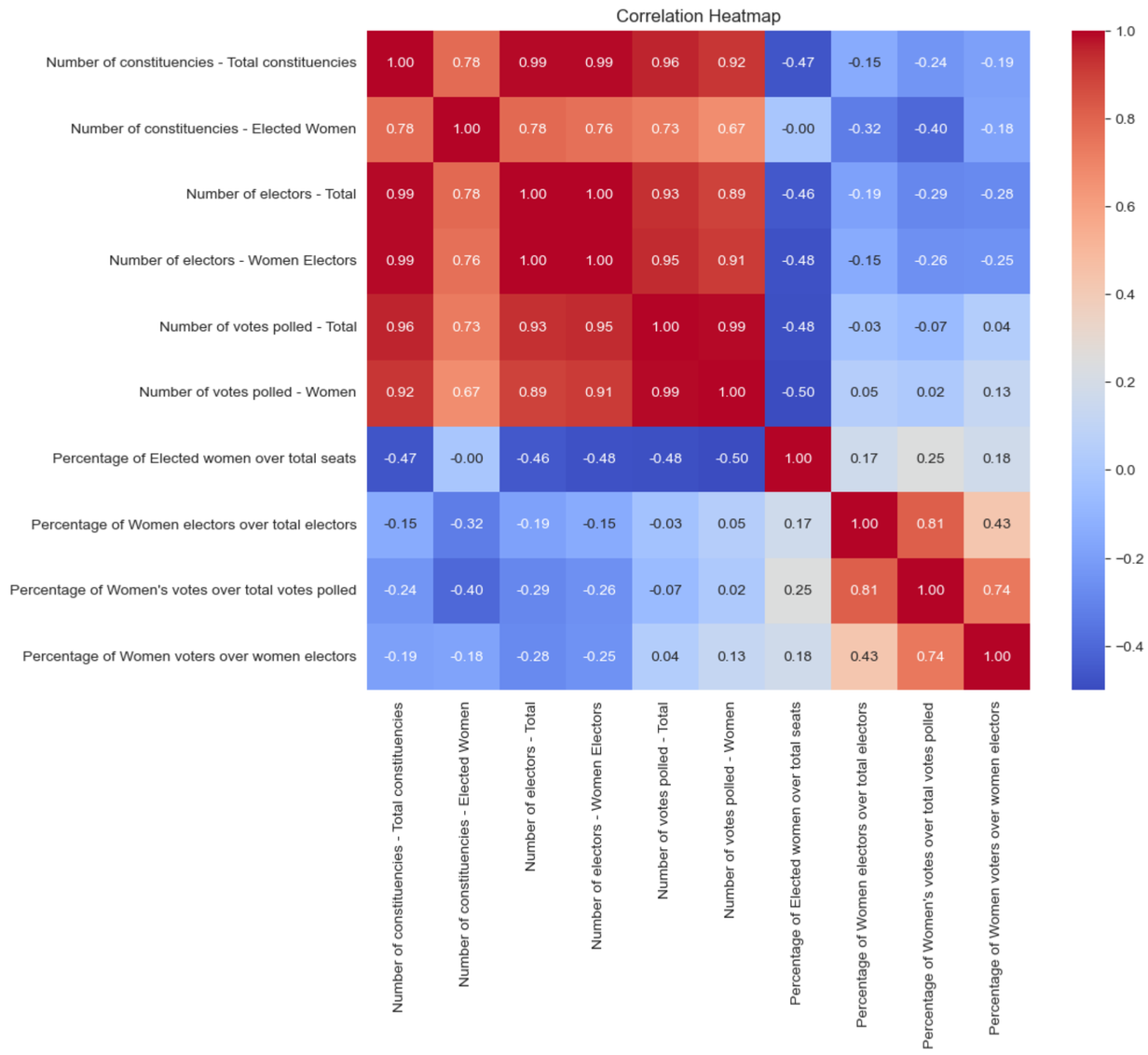
The distribution of total constituencies across states/UTs is skewed, with a majority of states having a lower number of constituencies. The total number of electors is also skewed, indicating that most states have a smaller electorate, while a few states have a very large electorate. The number of women electors follows a similar distribution to the total number of electors, suggesting that the proportion of women electors is consistent with the overall electorate distribution.

In [10]:
```
tqdm.pandas()

# Select only numeric columns and drop rows with NaN values
numeric_df = df.select_dtypes(include=["number"]).dropna()

# Calculate the correlation matrix
corr_matrix = numeric_df.corr()

# Generate a heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(corr_matrix, annot=True, fmt='.2f', cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```



INFERENCE:

the number of constituencies has a strong positive correlation with the total number of electors and the number of women electors, which suggests that larger constituencies tend to have more electors overall, including women.

Values close to -1 indicate a strong negative correlation, meaning as one variable increases, the other tends to decrease. The percentage of elected women over total seats has a moderately negative correlation with several variables such as the number of constituencies, total number of electors, and total votes polled, indicating that as these numbers increase, the proportion of elected women might decrease.

Values close to 0 indicate a weak or no correlation. For instance, the percentage of elected women over total seats has almost no correlation with the number of constituencies where women are elected, suggesting that the proportion of elected women does not depend on the number of constituencies where women are elected.

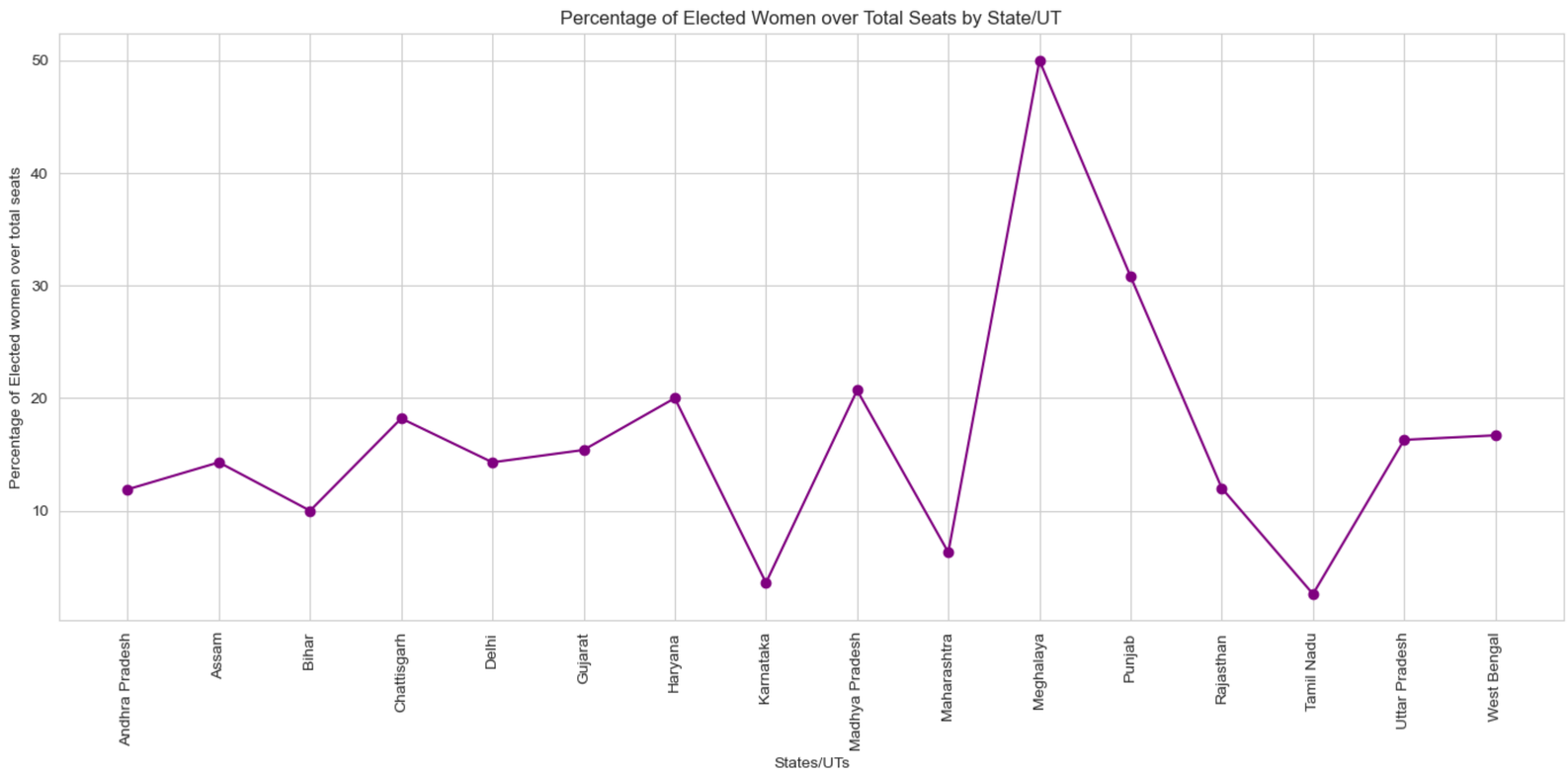
The diagonal from the top left to the bottom right shows perfect correlation (1.00) as it is comparing each variable to itself.

The heatmap shows that variables related to the number of electors and votes polled (total and women) are generally positively correlated with each other. In contrast, the percentages that compare women's figures to the total (e.g., percentage of elected women over total seats, percentage of women electors over total electors) tend to have negative correlations with these counts, which might suggest that while women's participation in terms of numbers is high, their representation proportionally is not as strong.

In [11]:
```
# Select relevant columns and drop rows with missing values
line_df = df[['States/UTs', 'Percentage of Elected women over total seats']].dropna()

# Sort the DataFrame by States/UTs
line_df.sort_values('States/UTs', inplace=True)

# Plot
plt.figure(figsize=(14, 7))
plt.plot(line_df['States/UTs'], line_df['Percentage of Elected women over total seats'], marker='o', linestyle='-', color='purple')
plt.xlabel('States/UTs')
plt.ylabel('Percentage of Elected women over total seats')
plt.xticks(rotation=90)
plt.title('Percentage of Elected Women over Total Seats by State/UT')
plt.tight_layout()
plt.savefig('percentage_elected_women_by_state.png')
plt.show()
```

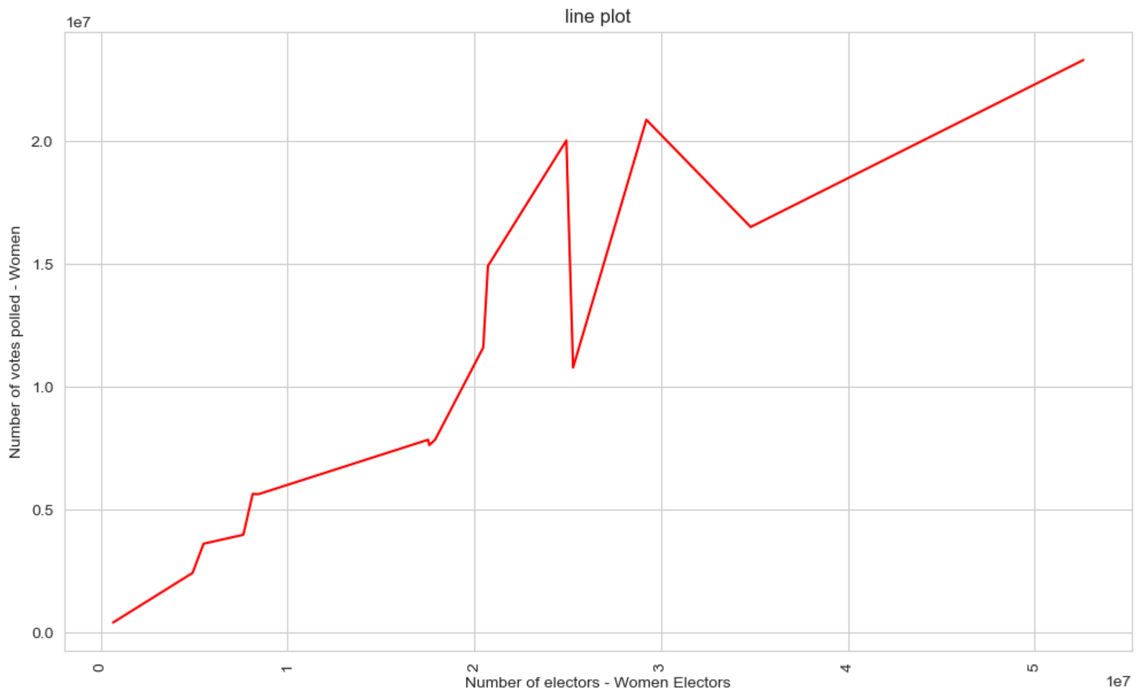


INFERENCE :

The percentage of elected women over total seats varies across states, with some states having a higher representation of women compared to others.

In [12]:
```
plt.figure(figsize = (12,7))
plt.title("line plot")
plt.xlabel('Number of electors - Women Electors')
plt.ylabel('Number of votes polled - Women')
plt.xticks(rotation=90)
sns.lineplot(data=df, x='Number of electors - Women Electors', y='Number of votes polled - Women', color='red' )
```

Out[12]:
```
<Axes: title={'center': 'line plot'}, xlabel='Number of electors - Women Electors', ylabel='Number of votes polled - Women'>
```



INFERENCE:

The line plot shows a general increasing trend with some fluctuations. There is a significant peak that exceeds 1.5 million votes, followed by a sharp decline and then a recovery back to the increasing trend.

Pie Chart of Percentage of Elected Women by State/UT:

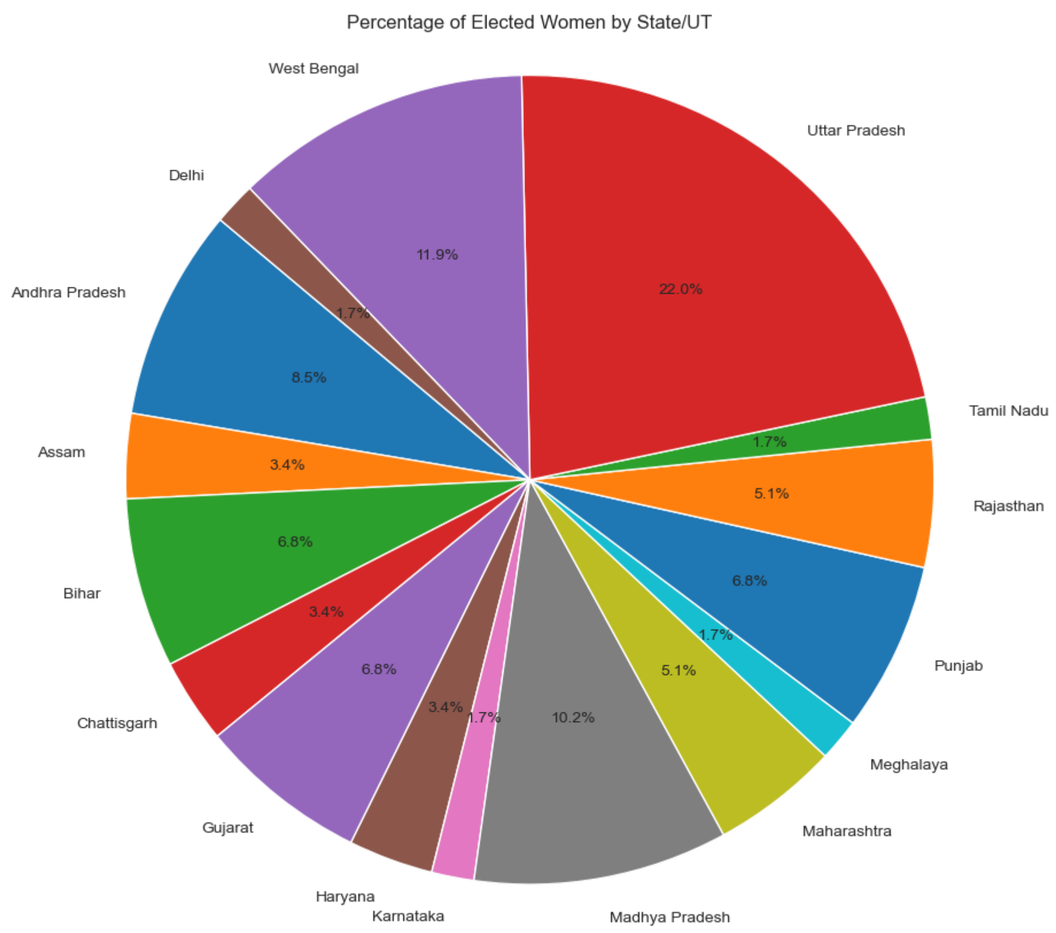
The pie chart shows the distribution of the percentage of elected women across different states, providing a visual comparison of the representation of women in different regions.

In [13]:
```
# Select relevant columns and drop rows with missing values
pie_df = df[['States/UTs', 'Number of constituencies - Elected Women']].dropna()

# Sum the total number of elected women
total_elected_women = pie_df['Number of constituencies - Elected Women'].sum()

# Calculate the percentage of elected women per state
pie_df['Percentage'] = (pie_df['Number of constituencies - Elected Women'] / total_elected_women) * 100

# Plot
plt.figure(figsize=(10, 10))
plt.pie(pie_df['Percentage'], labels=pie_df['States/UTs'], autopct='%1.1f%%', startangle=140)
plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.
plt.title('Percentage of Elected Women by State/UT')
plt.savefig('percentage_elected_women_pie_chart.png')
plt.show()
```



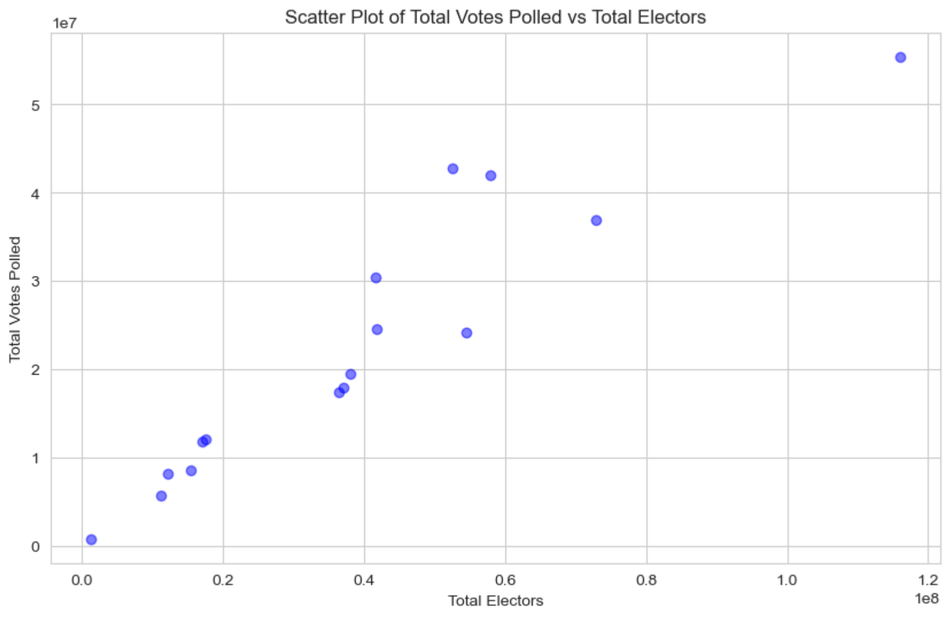
INFERENCE:

Uttar Pradesh has the largest segment with 22.0%. West Bengal follows with 11.9%. Madhya Pradesh has 10.2%. Other states and union territories like Tamil Nadu, Rajasthan, Punjab, Maharashtra, Karnataka, Haryana, Gujarat, Chhattisgarh, Bihar, Assam, Andhra Pradesh, and Delhi have smaller percentages ranging from 1.7% to 8.5%. The chart is a visual representation to compare the proportion of elected women across different regions, indicating where women's representation in elected positions is higher or lower.

In [14]:
```
import matplotlib.pyplot as plt

# Correcting the column names based on the dataframe
scatter_df = df[['Number of electors - Total', 'Number of votes polled - Total']].dropna()

# Plot
plt.figure(figsize=(10, 6))
plt.scatter(scatter_df['Number of electors - Total'], scatter_df['Number of votes polled - Total'], alpha=0.5, color='blue')
plt.xlabel('Total Electors')
plt.ylabel('Total Votes Polled')
plt.title('Scatter Plot of Total Votes Polled vs Total Electors')
plt.grid(True)
plt.savefig('scatter_total_votes_vs_electors_corrected.png')
plt.show()
```



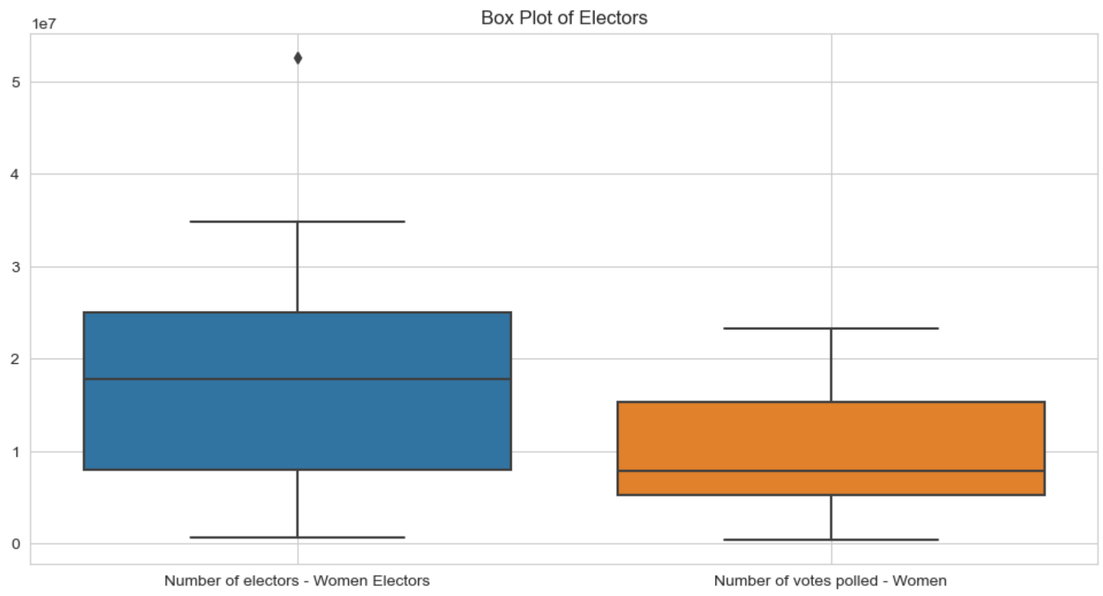
Scatter Plot of Total Votes Polled vs Total Electors:

The scatter plot shows the relationship between the total number of electors and the total votes polled, highlighting the distribution of data points.

In [15]:
```
# Select relevant columns and drop rows with missing values
box_df = df[['Number of electors - Women Electors', 'Number of votes polled - Women']].dropna()

# Plot
plt.figure(figsize=(12, 6))
sns.boxplot(data=box_df)
plt.xticks([0, 1], ['Number of electors - Women Electors', 'Number of votes polled - Women'])
plt.title('Box Plot of Electors')
plt.grid(True)
```

```
plt.savefig('boxplot_electors.png')
plt.show()
```



INFERENCE:

The first box plot on the left, colored in blue, represents the "Number of electors - Women Electors". The second box plot on the right, colored in orange, represents the "Number of votes polled - Women". Both box plots are plotted on the same y-axis, which is scaled from 0 to 1e7 (10 million), indicating the count of electors or votes. The x-axis labels the two categories being compared.

Here are some observations from the box plots:

The range of the number of electors (blue box plot) is narrower than the range of the number of votes polled (orange box plot). The median value for both categories is above 1e7/2 (5 million), with the median for the number of electors appearing to be slightly higher than the median for the number of votes polled. There are outliers in the number of electors as indicated by the diamond shape above the upper whisker of the blue box plot.

SUMMARY:

The dataset provides information on the political representation and participation of women in various states and union territories of a country. It includes data on:

Total number of constituencies Number of elected women representatives Total number of electors Number of women electors Total votes polled Votes polled by women Various percentages related to women's participation and representation in the electoral process The preview shows data for Andhra Pradesh, Arunachal Pradesh, and Assam, with various metrics such as the number of constituencies, electors, votes polled, and percentages of women's participation.

REFERENCE

https://data.gov.in/