

```
In [3]: import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns
```

```
In [4]: df_2001=pd.read_csv("rape_2001.csv")  
df_2011=pd.read_csv("rape_2011.csv")
```

```
In [5]: df_2011.head()
```

Out[5]:

Category	States/ UTs/ Cities	No. of Incest (Rape) Cases Reported	No. of Victims of Incest (Rape)	No. of Victims of Incest (Rape) - Upto 10 Years	No. of Victims of Incest (Rape) - 10-14 Years	No. of Victims of Incest (Rape) - 14-18 Years	No. of Victims of Incest (Rape) - 18-30 Years	No. of Victims of Incest (Rape) - 30-50 Years	No. of Victims of Incest (Rape) - Above 50 Years	No. of Victims of Incest (Rape) - Total	No. of Victims of Incest (Rape)	No. of Victims of Incest (Rape)	
0	State	Andhra Pradesh	0	0	0	0	0	0	0	0	0	0	...
1	State	Arunachal Pradesh	0	0	0	0	0	0	0	0	0	0	...
2	State	Assam	0	0	0	0	0	0	0	0	0	0	...
3	State	Bihar	2	2	0	0	0	0	0	0	0	2	...
4	State	Chhattisgarh	39	0	3	31	4	1	0	39	39	39	...

5 rows × 26 columns

```
In [6]: df_2001.head()
```

Out[6]:

Category	States/UTs/Cities	Reported (Incest Rape Cases)	No. of 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
			10 Years	14 Years	18 Years	30 Years	50 Years	Years		
0	State	Andhra Pradesh	4.0	0.0	2.0	0.0	2.0	0.0	0.0	4.0
1	State	Arunachal Pradesh	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	State	Assam	8.0	0.0	1.0	7.0	0.0	0.0	0.0	8.0
3	State	Bihar	8.0	0.0	2.0	5.0	1.0	0.0	0.0	8.0
4	State	Chhattisgarh	6.0	0.0	3.0	1.0	1.0	1.0	0.0	6.0

5 rows × 26 columns

In [7]:

```
df_2001['year']=2001
df_2011['year']=2011
```

In [8]:

```
# The standardized column names from the 2001 dataset. this is done because the two dataframes have different column names
standard_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"
```

]

```
# Option 1: Assign these column names to just the relevant portion
# (if 2011 dataset has these columns at the end, e.g., columns 2 to 25)
df_2011.columns.values[2:26] = standard_columns
```

```
# OR Option 2: If the entire df_2011 has only these 24 columns in correct order
```

```
# df_2011.columns = standard_columns  
  
# Verify  
print(df_2011.columns[2:26])
```

```
Index(['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'],  
      dtype='object')
```

```
In [9]: df_2011.head()
```

Out[9]:

Category	States/ UTs/ Cities	Reported (Incest Rape Cases)	No. of Cases	No. of (Incest Rape Cases)								
			upto 10 Years	- 10 - 14 Years	- 14 - 18 Years	- 18 - 30 Years	- 30 - 50 Years	- 50 - 50 Years	- 50 - Total Victims			
0	State	Andhra Pradesh	0	0	0	0	0	0	0	0	0	0
1	State	Arunachal Pradesh	0	0	0	0	0	0	0	0	0	0
2	State	Assam	0	0	0	0	0	0	0	0	0	0
3	State	Bihar	2	2	0	0	0	0	0	0	0	2
4	State	Chhattisgarh	39	0	3	31	4	1	0	39	39	...

5 rows \times 27 columns

```
In [10]: df_2001.head()
```

Out[10]:

			No. of Cases Reported (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
0	State	Andhra Pradesh	4.0	0.0	2.0	0.0	2.0	0.0	0.0
1	State	Arunachal Pradesh	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	State	Assam	8.0	0.0	1.0	7.0	0.0	0.0	8.0
3	State	Bihar	8.0	0.0	2.0	5.0	1.0	0.0	8.0
4	State	Chhattisgarh	6.0	0.0	3.0	1.0	1.0	1.0	6.0

5 rows × 27 columns

In [14]: `df_2011.rename(columns={'States/ UTs/ Cities': 'States/UTs/Cities'}, inplace=True)`

In [15]: `combined_df = pd.concat([df_2001, df_2011], ignore_index=True)`
`combined_df.head()`

Out[15]:

			No. of Cases Reported (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
0	State	Andhra Pradesh	4.0	0.0	2.0	0.0	2.0	0.0	0.0
1	State	Arunachal Pradesh	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	State	Assam	8.0	0.0	1.0	7.0	0.0	0.0	8.0
3	State	Bihar	8.0	0.0	2.0	5.0	1.0	0.0	8.0
4	State	Chhattisgarh	6.0	0.0	3.0	1.0	1.0	1.0	6.0

5 rows × 27 columns

In [16]: `combined_df['year'].unique`

```
Out[16]: <bound method Series.unique of 0      2001  
          1      2001  
          2      2001  
          3      2001  
          4      2001  
          ...  
         161    2011  
         162    2011  
         163    2011  
         164    2011  
         165    2011  
Name: year, Length: 166, dtype: int64>
```

```
In [19]: combined df.describe()
```

Out[19] :

	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
count	163.000000	163.000000	163.000000	163.000000	163.000000	163.000000	163.000000	163.000000
mean	13.926380	0.920245	2.300613	4.233129	4.705521	1.748466	0.018405	13.926380
std	59.762551	3.488927	8.916025	17.114563	22.499090	9.059708	0.134825	59.762551
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
75%	4.000000	0.000000	0.000000	1.000000	1.000000	0.000000	0.000000	4.000000
max	439.000000	24.000000	63.000000	112.000000	170.000000	69.000000	1.000000	439.000000

8 rows × 25 columns

In [108...]

```
# Filter rows where "Total Rape Cases" column has NA or 0 values
na_or_zero_rows = combined_df[(combined_df['No. of Cases Reported (Total Rape Cases)'] == combined_df['No. of Cases Reported (Total Rape Cases)']) == 0]

# Display the corresponding state/UT/city and year
print(na_or_zero_rows[['States/UTs/Cities', 'year']])
```

States/UTs/Cities	year
32 Daman & Diu	2001
34 Lakshadweep	2001
42 Asansol	2001
48 Dhanbad	2001
54 Jamshedpur	2001
108 Lakshadweep	2011

In [112...]

```
cols=['Daman & Diu', 'Asansol', 'Dhanbad', 'Jamshedpur']
combined_df[(combined_df['States/UTs/Cities'].isin(cols)) & (combined_df['year'] == 2011)]
```

Out[112]:

Category	States/UTs/Cities	Reported (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 Victi
			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 Victi
106	Union Territory	Daman & Diu	0.0	0.0	0.0	0.0	0.0	0.0
116	City	Asansol	0.0	0.0	0.0	0.0	0.0	0.0
124	City	Dhanbad	0.0	0.0	0.0	0.0	0.0	0.0
133	City	Jamshedpur	0.0	0.0	0.0	0.0	0.0	0.0

4 rows × 27 columns

INTERPRETATION: these four cities namely 'Daman & Diu','Asansol','Dhanbad','Jamshedpur' have seen other rape cases in 2011 when they didn't had any rape cases in 2001, especially asansol which saw 31 other rape cases in the year 2011.

```
In [23]: # List of labels to exclude from 'States/UTs/Cities' column
excluded_labels = ["Total (States)", "Total (UTs)", "Total (Cities)", "Total (All-India)"]

# Filter out these rows from the combined DataFrame
filtered_df = combined_df[~combined_df["States/UTs/Cities"].isin(excluded_labels)]

# For 2001
max_2001 = filtered_df[filtered_df['year'] == 2001]["No. of Cases Reported (Total Rape Cases)"]
print("Max rape cases in 2001 (excluding totals):", max_2001)

# For 2011
max_2011 = filtered_df[filtered_df['year'] == 2011]["No. of Cases Reported (Total Rape Cases)"]
print("Max rape cases in 2011 (excluding totals):", max_2011)
```

Max rape cases in 2001 (excluding totals): 2851.0

Max rape cases in 2011 (excluding totals): 3406.0

```
In [ ]: # Filter rows where "Total Rape Cases" column has NA values
na_rows_total_rape_cases = combined_df[combined_df['No. of Cases Reported (Total Rape Cases)'].notna()]

# Display the corresponding state/UT/city and year
print(na_rows_total_rape_cases[['States/UTs/Cities', 'year']])
```

```
In [63]: # state/ut/city with highest number of total rape cases for both the years

max_rows = filtered_df.loc[filtered_df.groupby("year")["No. of Cases Reported (Total Rape Cases)"].idxmax()]
print(max_rows[['States/UTs/Cities', 'year', 'No. of Cases Reported (Total Rape Cases)'])
```

States/UTs/Cities	year	No. of Cases Reported (Total Rape Cases)	
13	Madhya Pradesh	2001	2851.0
87	Madhya Pradesh	2011	3406.0

NOTE: MADHYA PRADESH HAS THE HIGHEST NUMBER OF RAPE CASES FOR BOTH THE YEARS.

```
In [59]: #top 5 states/uts/cities with highest reported rape cases(total rape cases) for both t
sorted_df = filtered_df.sort_values(by=["year", "No. of Cases Reported (Total Rape Cas
top_5_cases_per_year = sorted_df.groupby("year").head(5)

print(top_5_cases_per_year[top_5_cases_per_year['year'] == 2001][['States/UTs/Cities', 'y
print(top_5_cases_per_year[top_5_cases_per_year['year'] == 2011][['States/UTs/Cities', 'y

      States/UTs/Cities  year  No. of Cases Reported (Total Rape Cases)
13    Madhya Pradesh  2001           2851.0
25    Uttar Pradesh  2001           1958.0
14    Maharashtra  2001           1302.0
21    Rajasthan  2001           1049.0
4     Chhattisgarh  2001            959.0
      States/UTs/Cities  year  No. of Cases Reported (Total Rape Cases)
87    Madhya Pradesh  2011           3406.0
101   West Bengal  2011           2363.0
99    Uttar Pradesh  2011           2042.0
95    Rajasthan  2011           1800.0
88    Maharashtra  2011           1701.0
```

```
In [51]: #states/ut's/citie's with change in the rape cases from 2001 to 2011
grouped_df = filtered_df.groupby(['States/UTs/Cities', 'year'])['No. of Cases Reported

data_2001 = grouped_df[grouped_df['year'] == 2001].rename(columns={'No. of Cases Repor
data_2011 = grouped_df[grouped_df['year'] == 2011].rename(columns={'No. of Cases Repor

merged = pd.merge(data_2001[['States/UTs/Cities', 'cases_2001']], data_2011[['States/U

merged['change'] = merged['cases_2011'] - merged['cases_2001']

print(merged[['States/UTs/Cities', 'cases_2001', 'cases_2011', 'change']])
```

	States/UTs/Cities	cases_2001	cases_2011	change
0	A & N Islands	3.0	13.0	10.0
1	Agra	66.0	61.0	-5.0
2	Ahmedabad	39.0	60.0	21.0
3	Allahabad	7.0	10.0	3.0
4	Amritsar	7.0	20.0	13.0
..
60	Vadodara	3.0	9.0	6.0
61	Varanasi	17.0	5.0	-12.0
62	Vijayawada	25.0	37.0	12.0
63	Vishakhapatnam	13.0	57.0	44.0
64	West Bengal	709.0	2363.0	1654.0

[65 rows x 4 columns]

```
In [62]: #states/ut's/cities with decrease in rape cases from 2001 to 2011
decreased_cases = merged[merged['change'] < 0]
print(decreased_cases[['States/UTs/Cities', 'cases_2001', 'cases_2011', 'change']])
```

	States/UTs/Cities	cases_2001	cases_2011	change
1	Agra	66.0	61.0	-5.0
15	D & N Haveli	6.0	4.0	-2.0
37	Lucknow	39.0	38.0	-1.0
61	Varanasi	17.0	5.0	-12.0

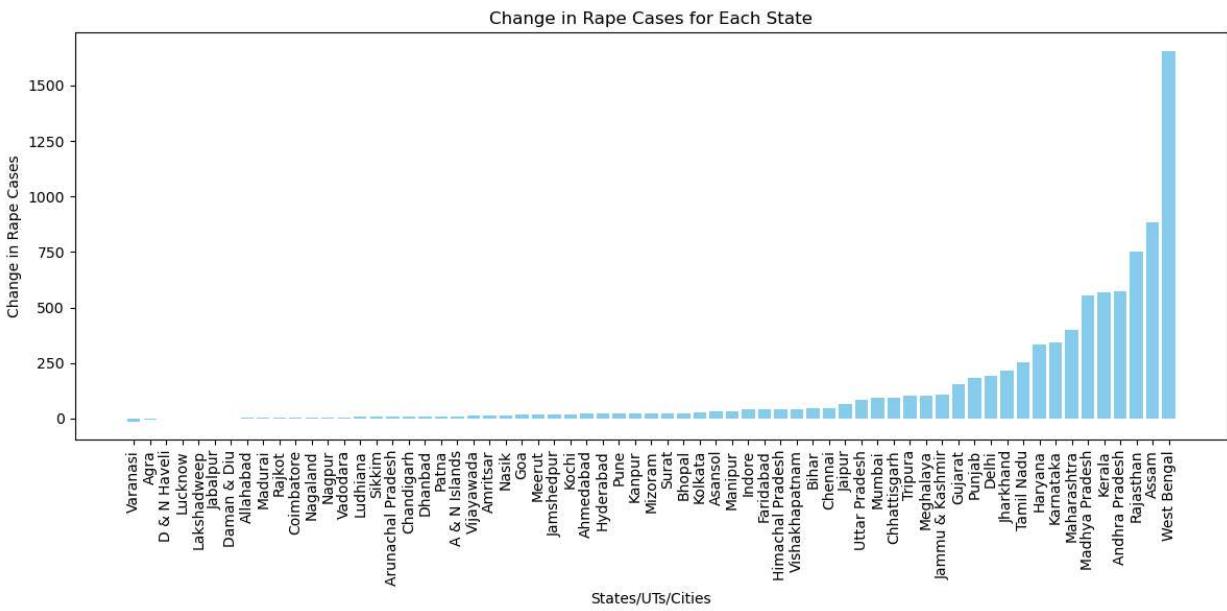
NOTE: NONE OF THE STATE/UT AS A WHOLE HAS DECREASE IN RAPE CASES, HOWEVER THERE ARE FOUR CITIES WITH MINIMAL DECREASE IN RAPE CASES.

```
In [53]: #state/ut's/cities with maximum increase in rape case from 2001 to 2011  
state_with_maximum_increase= merged.loc[merged['change'].idxmax(),['States/UTs/Cities']]  
print(state_with_maximum_increase)
```

```
States/UTs/Cities      West Bengal  
change                  1654.0  
Name: 64, dtype: object
```

NOTE: WEST BENGAL HAS THE HIGHEST INCREASE IN TOTAL RAPE CASES FROM 2001 TO 2011

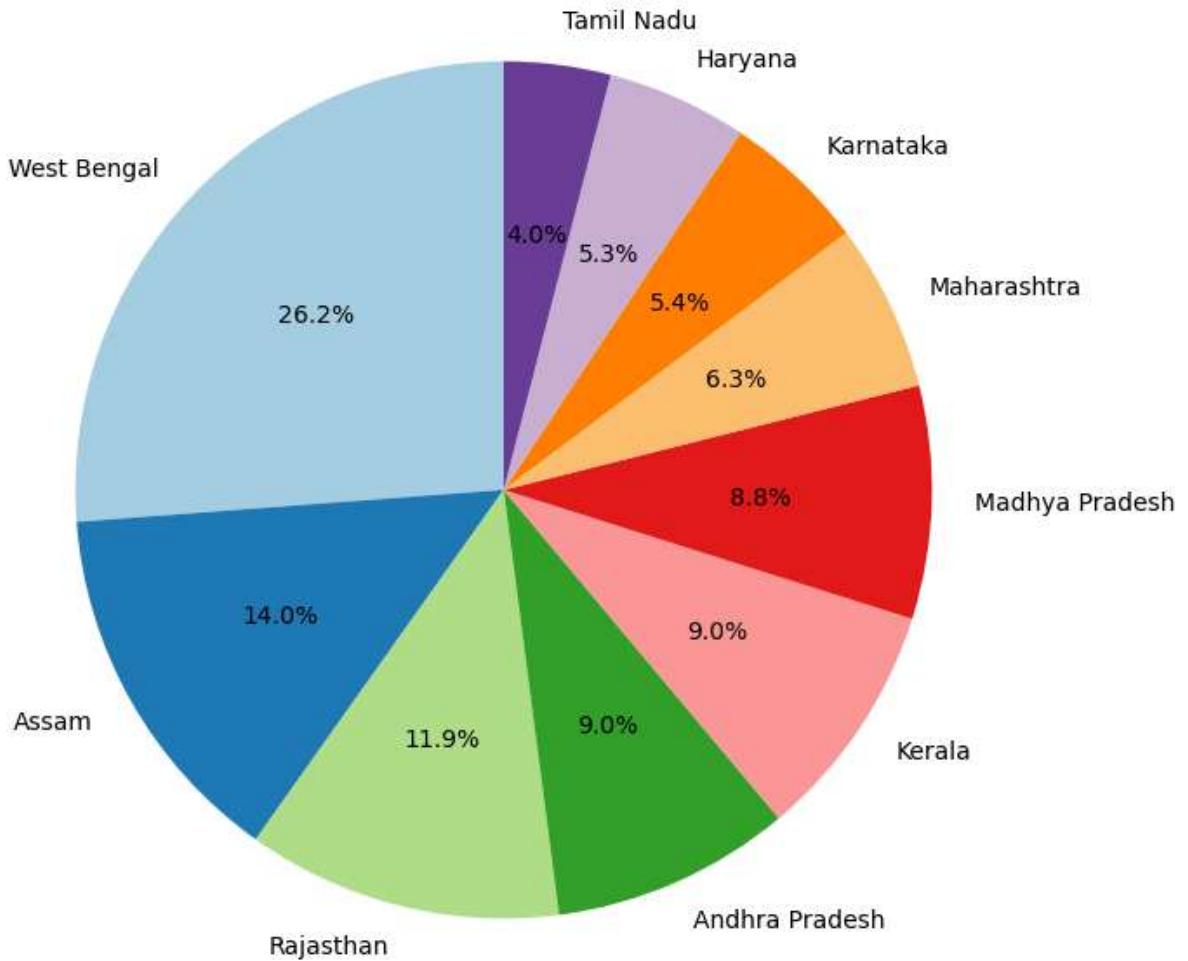
```
In [66]: #bar plot showing change in total rape cases for each state from 2001 to 2011
sorted_data=merged.sort_values(by='change',ascending=True)
plt.figure(figsize=(12, 6))
plt.bar(sorted_data['States/UTs/Cities'], sorted_data['change'], color='skyblue')
plt.xlabel('States/UTs/Cities')
plt.ylabel('Change in Rape Cases')
plt.title('Change in Rape Cases for Each State')
plt.xticks(rotation=90) # Rotate state names for better readability
plt.tight_layout()
plt.show()
```



```
In [65]: # You might want to focus on the top N states with the largest changes
top_states = merged.nlargest(10, 'change') # Get top 10 states with the highest change

plt.figure(figsize=(8, 8))
plt.pie(top_states['change'], labels=top_states['States/UTs/Cities'], autopct='%1.1f%%')
plt.title('Top 10 States with Change in Rape Cases')
plt.show()
```

Top 10 States with Change in Rape Cases



```
In [72]: #change in total rape cases of total(all india),states,uts and cities
categories = ["Total (States)", "Total (UTs)", "Total (Cities)",'Total (All-India)']

for category in categories:
    total_rape_2001 = combined_df[(combined_df['States/UTs/Cities'] == category) & (combined_df['Year'] == 2001)]
    total_rape_2011 = combined_df[(combined_df['States/UTs/Cities'] == category) & (combined_df['Year'] == 2011)]
    change = total_rape_2011 - total_rape_2001
    print(f"Change in total rape cases for {category} between 2001 and 2011: {change}")


Change in total rape cases for Total (States) between 2001 and 2011: 7924.0
Change in total rape cases for Total (UTs) between 2001 and 2011: 207.0
Change in total rape cases for Total (Cities) between 2001 and 2011: 1318.0
Change in total rape cases for Total (All-India) between 2001 and 2011: 8131.0
```

```
In [77]: max_rows = filtered_df.loc[filtered_df.groupby("year")["No. of Victims (Incest Rape Cases)"].idxmax()]
print(max_rows[["States/UTs/Cities", "year", "No. of Victims (Incest Rape Cases)"]]

   States/UTs/Cities  year  No. of Victims (Incest Rape Cases)
13      Madhya Pradesh  2001                      10.0
86        Kerala        2011                       6.0
```

```
In [87]: for col in standard_columns:
    max_rows=filtered_df.loc[filtered_df.groupby("year")[col].idxmax()]
```

```
print("State With Maximum ",col,"For Each Year:")
print(max_rows[["States/UTs/Cities", "year", col]])
```

```
print()
```

State With Maximum No. of Cases Reported (Incest Rape Cases) For Each Year:

States/UTs/Cities	year	No. of Cases Reported (Incest Rape Cases)
13	Madhya Pradesh	2001
88	Maharashtra	2011

State With Maximum No. of Victims (Incest Rape Cases) upto 10 Years For Each Year:

States/UTs/Cities	year	No. of Victims (Incest Rape Cases) upto 10 Years
13	Madhya Pradesh	2001
86	Kerala	2011

State With Maximum No. of Victims (Incest Rape Cases) - 10 - 14 Years For Each Year:

States/UTs/Cities	year	No. of Victims (Incest Rape Cases) - 10 - 14 Years
13	Madhya Pradesh	2001
86	Kerala	2011

State With Maximum No. of Victims (Incest Rape Cases) - 14 - 18 Years For Each Year:

States/UTs/Cities	year	No. of Victims (Incest Rape Cases) - 14 - 18 Years
13	Madhya Pradesh	2001
78	Chhattisgarh	2011

State With Maximum No. of Victims (Incest Rape Cases) - 18 - 30 Years For Each Year:

States/UTs/Cities	year	No. of Victims (Incest Rape Cases) - 18 - 30 Years
13	Madhya Pradesh	2001
95	Rajasthan	2011

State With Maximum No. of Victims (Incest Rape Cases) - 30 - 50 Years For Each Year:

States/UTs/Cities	year	No. of Victims (Incest Rape Cases) - 30 - 50 Years
13	Madhya Pradesh	2001
95	Rajasthan	2011

State With Maximum No. of Victims (Incest Rape Cases) above 50 Years For Each Year:

States/UTs/Cities	year	No. of Victims (Incest Rape Cases) above 50 Years
12	Kerala	2001
74	Andhra Pradesh	2011

State With Maximum No. of Victims (Incest Rape Cases) - Total Victims For Each Year:

States/UTs/Cities	year	No. of Victims (Incest Rape Cases) - Total Victims
13	Madhya Pradesh	2001
88	Maharashtra	2011

State With Maximum No. of Cases Reported (Other Rape Cases) For Each Year:

States/UTs/Cities	year	No. of Cases Reported (Other Rape Cases)
13	Madhya Pradesh	2001
87	Madhya Pradesh	2011

State With Maximum No. of Victims (Other Rape Cases) upto 10 Years For Each Year:

States/UTs/Cities	year	No. of Victims (Other Rape Cases) upto 10 Years
14	Maharashtra	2001
99	Uttar Pradesh	2011

State With Maximum No. of Victims (Other Rape Cases) - 10 - 14 Years For Each Year:

States/UTs/Cities	year	No. of Victims (Other Rape Cases) - 10 - 14 Years
13	Madhya Pradesh	2001
87	Madhya Pradesh	2011

State With Maximum No. of Victims (Other Rape Cases) - 14 - 18 Years For Each Year:

States/UTs/Cities	year	No. of Victims (Other Rape Cases) - 14 - 18 Years
13	Madhya Pradesh	2001
87	Madhya Pradesh	2011

State With Maximum No. of Victims (Other Rape Cases) - 18 - 30 Years For Each Year:

States/UTs/Cities		year	No. of Victims (Other Rape Cases) - 18 - 30 Years
13	Madhya Pradesh	2001	1120.0
101	West Bengal	2011	1671.0

State With Maximum No. of Victims (Other Rape Cases) - 30 - 50 Years For Each Year:

States/UTs/Cities		year	No. of Victims (Other Rape Cases) - 30 - 50 Years
13	Madhya Pradesh	2001	571.0
87	Madhya Pradesh	2011	546.0

State With Maximum No. of Victims (Other Rape Cases) above 50 Years For Each Year:

States/UTs/Cities		year	No. of Victims (Other Rape Cases) above 50 Years
13	Madhya Pradesh	2001	27.0
87	Madhya Pradesh	2011	27.0

State With Maximum No. of Victims (Other Rape Cases) - Total Victims For Each Year:

States/UTs/Cities		year	No. of Victims (Other Rape Cases) - Total Victims
13	Madhya Pradesh	2001	2550.0
87	Madhya Pradesh	2011	3398.0

State With Maximum No. of Cases Reported (Total Rape Cases) For Each Year:

States/UTs/Cities		year	No. of Cases Reported (Total Rape Cases)
13	Madhya Pradesh	2001	2851.0
87	Madhya Pradesh	2011	3406.0

State With Maximum No. of Victims (Total Rape Cases) upto 10 Years For Each Year:

States/UTs/Cities		year	No. of Victims (Total Rape Cases) upto 10 Years
14	Maharashtra	2001	89.0
99	Uttar Pradesh	2011	150.0

State With Maximum No. of Victims (Total Rape Cases) - 10-14 Years For Each Year:

States/UTs/Cities		year	No. of Victims (Total Rape Cases) - 10-14 Years
13	Madhya Pradesh	2001	304.0
87	Madhya Pradesh	2011	309.0

State With Maximum No. of Victims (Total Rape Cases) - 14 - 18 Years For Each Year:

States/UTs/Cities		year	No. of Victims (Total Rape Cases) - 14 - 18 Years
13	Madhya Pradesh	2001	557.0
87	Madhya Pradesh	2011	886.0

State With Maximum No. of Victims (Total Rape Cases) - 18 - 30 Years For Each Year:

States/UTs/Cities		year	No. of Victims (Total Rape Cases) - 18 - 30 Years
13	Madhya Pradesh	2001	1250.0
101	West Bengal	2011	1671.0

State With Maximum No. of Victims (Total Rape Cases) - 30 - 50 Years For Each Year:

States/UTs/Cities		year	No. of Victims (Total Rape Cases) - 30 - 50 Years
13	Madhya Pradesh	2001	627.0
87	Madhya Pradesh	2011	546.0

State With Maximum No. of Victims (Total Rape Cases) above 50 Years For Each Year:

States/UTs/Cities		year	No. of Victims (Total Rape Cases) above 50 Years
13	Madhya Pradesh	2001	27.0
87	Madhya Pradesh	2011	27.0

State With Maximum No. of Victims (Total Rape Cases) - Total Victims For Each Year:

States/UTs/Cities		year	No. of Victims (Total Rape Cases) - Total Victims
13	Madhya Pradesh	2001	2851.0
87	Madhya Pradesh	2011	3408.0

```
In [89]: filtered_df.to_excel("rape_analysis.xlsx", index=False)
```

```
In [90]: from io import BytesIO
from IPython.display import FileLink

# Create a BytesIO buffer
output = BytesIO()

# Write the DataFrame to the buffer
filtered_df.to_excel(output, index=False)

# Save the buffer content to a downloadable Link
with open("output.xlsx", "wb") as f:
    f.write(output.getbuffer())

# Display download link
FileLink("output.xlsx")
```

```
Out[90]: output.xlsx
```

```
In [98]: india_df = combined_df[combined_df['States/UTs/Cities'] == 'Total (All-India)']

# Specify your age group columns
age_group_cols = ['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']
label=['upto 10 years','10-14 years','14-18 years','18-30 years','30-50 years','above

# Separate rows for each year
data_2001 = india_df[india_df['year'] == 2001][age_group_cols].iloc[0]
data_2011 = india_df[india_df['year'] == 2011][age_group_cols].iloc[0]

fig, axs = plt.subplots(2, 1, figsize=(8, 12)) # 2 rows, 1 column (vertically stacked

# Smaller font for labels
label_font = {'fontsize': 10} # Adjust this as needed

# Pie for 2001
axs[0].pie(
    data_2001,
    labels=label,
    autopct='%1.1f%%',
    startangle=90,
    textprops={'fontsize': 12},
    labeldistance=1.1 # Adjusts label position slightly outside the pie
)
axs[0].set_title("Incest Rape by Age Group (2001)", fontsize=14)

# Reduce only the label font size
for text in axs[0].texts[:len(age_group_cols)]:
    text.set_fontsize(label_font['fontsize'])

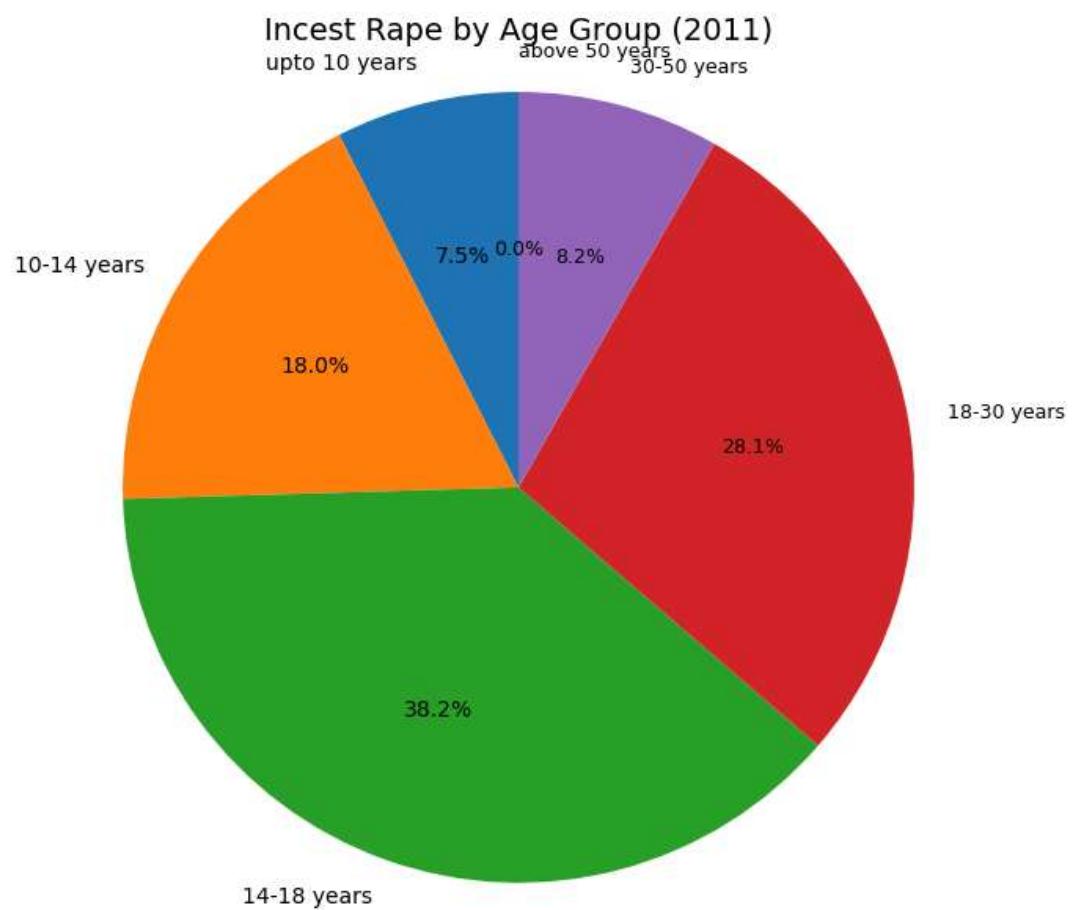
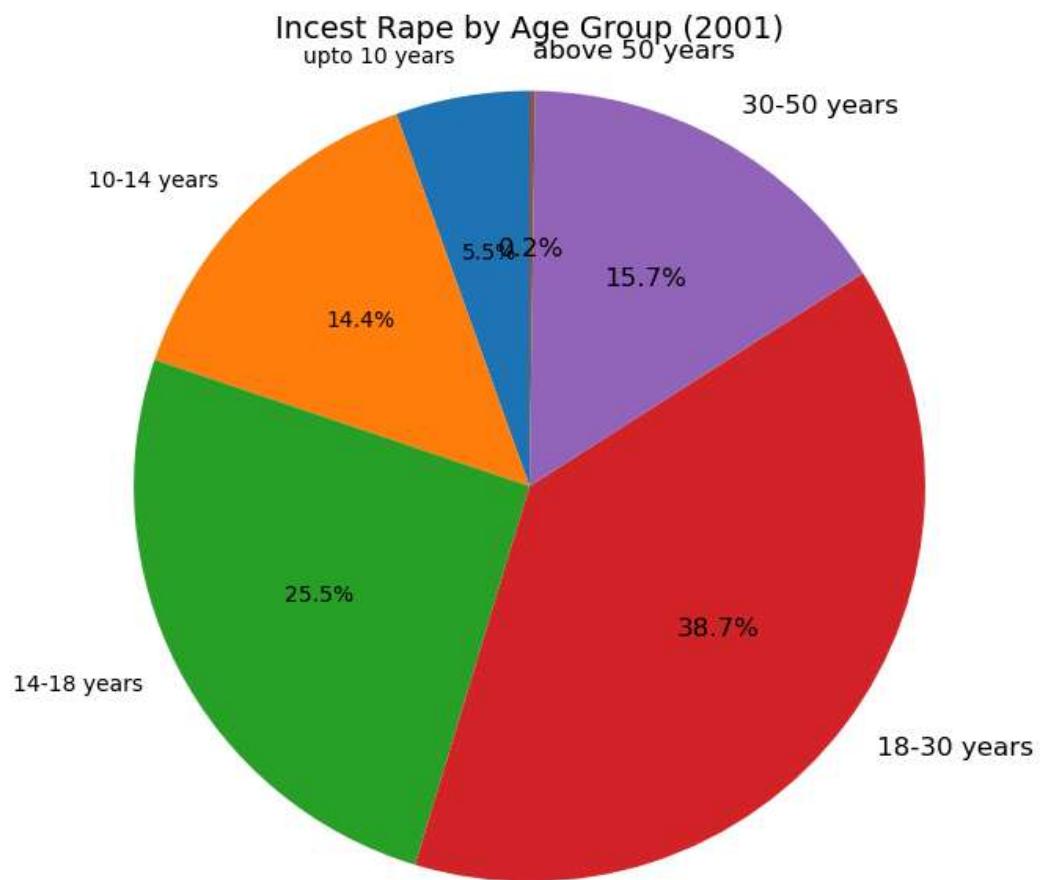
# Pie for 2011
axs[1].pie(
    data_2011,
```

```
    labels=label,
    autopct='%1.1f%%',
    startangle=90,
    textprops={'fontsize': 9},
    labeldistance=1.1
)
axs[1].set_title("Incest Rape by Age Group (2011)", fontsize=14)

# Reduce Label font size
for text in axs[1].texts[:len(age_group_cols)]:
    text.set_fontsize(label_font['fontsize'])

# Equal aspect ratio makes pie look like a circle
for ax in axs:
    ax.axis('equal')

plt.tight_layout() # Ensure everything fits without overlap
plt.show()
```



INTERPRETATION: IN 2001 THE AGE GROUP OF 18-30 YEARS HAD THE HIGHEST PROPORTION OF INCEST RAPE CASES(38.7%) BUT IN THE YEAR 2011 THE AGE GROUP OF 14-18 YEARS HAD THE HIGHEST PROPORTION OF INCEST RAPE CASES(38.2%), ALSO THERE HAS BEEN AN INCREASE IN THE INCEST RAPE CASES OF YOUNGER AGE GROUPS WHICH SHOWS THAT 2011 WHEN COMPARED TO 2001 , INDIA WITNESSED MORE CHILD/MINOR INCEST RAPE CASES WHICH IS VERY DISTURBING AND CONCERNING.

```
In [101...]: agecols=["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"]

data1_2001 = india_df[india_df['year'] == 2001][agecols].iloc[0]
data1_2011 = india_df[india_df['year'] == 2011][agecols].iloc[0]

fig, axs = plt.subplots(2, 1, figsize=(8, 12)) # 2 rows, 1 column (vertically stacked)

# Smaller font for labels
label_font = {'fontsize': 10} # Adjust this as needed

# Pie for 2001
axs[0].pie(
    data1_2001,
    labels=label,
    autopct='%1.1f%%',
    startangle=90,
    textprops={'fontsize': 12},
    labeldistance=1.1 # Adjusts label position slightly outside the pie
)
axs[0].set_title("Total Rape Victims by Age Group (2001)", fontsize=14)

# Reduce only the label font size
for text in axs[0].texts[:len(age_group_cols)]:
    text.set_fontsize(label_font['fontsize'])

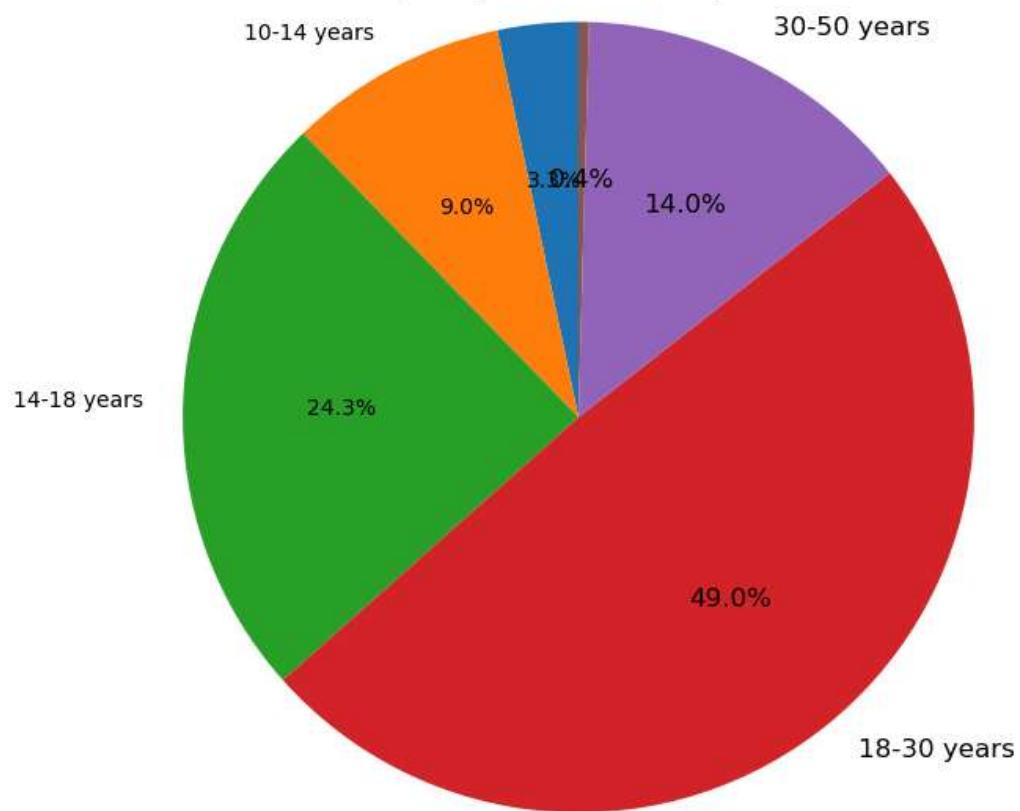
# Pie for 2011
axs[1].pie(
    data1_2011,
    labels=label,
    autopct='%1.1f%%',
    startangle=90,
    textprops={'fontsize': 9},
    labeldistance=1.1
)
axs[1].set_title("Incest Rape by Age Group (2011)", fontsize=14)

# Reduce Label font size
for text in axs[1].texts[:len(age_group_cols)]:
    text.set_fontsize(label_font['fontsize'])

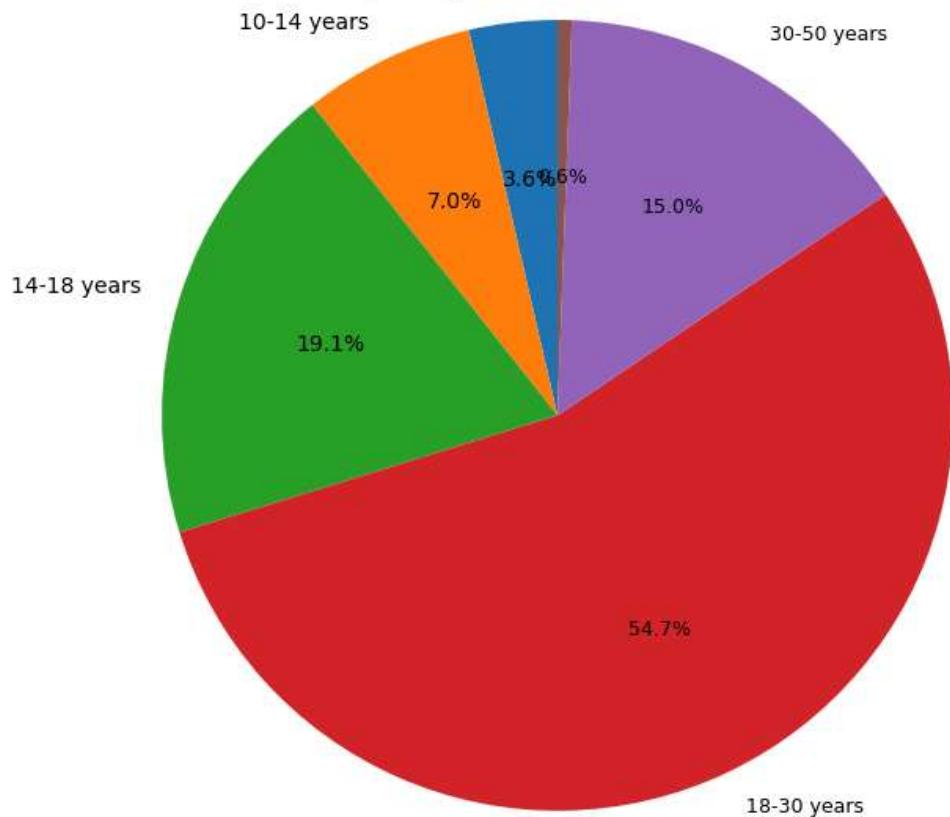
# Equal aspect ratio makes pie look like a circle
for ax in axs:
    ax.axis('equal')
```

```
plt.tight_layout() # Ensure everything fits without overlap  
plt.show()
```

Total Rape Victims by Age Group (2001)



Incest Rape by Age Group (2011)



NOTE: WHEN WE LOOK AT THE PIE CHARTS OF THESE TWO YEARS SHOWING TOTAL RAPE CASES WE DONT SEE ANY INCREASE IN CASES OF MINOR RAPE CASES WHICH MEANS THAT THE MAJORITY OF CHILD RAPE CASES ARE INCEST RAPE CASES.

In [114...]

```
import matplotlib.pyplot as plt

# Calculate the percentage of incest rape cases to total rape cases for each year and
filtered_df['Incest to Total Rape Cases (%)'] = (filtered_df['No. of Victims (Incest R
ape)'] / filtered_df['Total Rape Cases']) * 100

# Pivot the data to get the percentage for each state and year
pivot_df = filtered_df.pivot_table(values='Incest to Total Rape Cases (%)', index='Sta
te', columns='Year', aggfunc='mean')

# Plot stacked bar chart
pivot_df.plot(kind='bar', stacked=True, figsize=(12, 7))

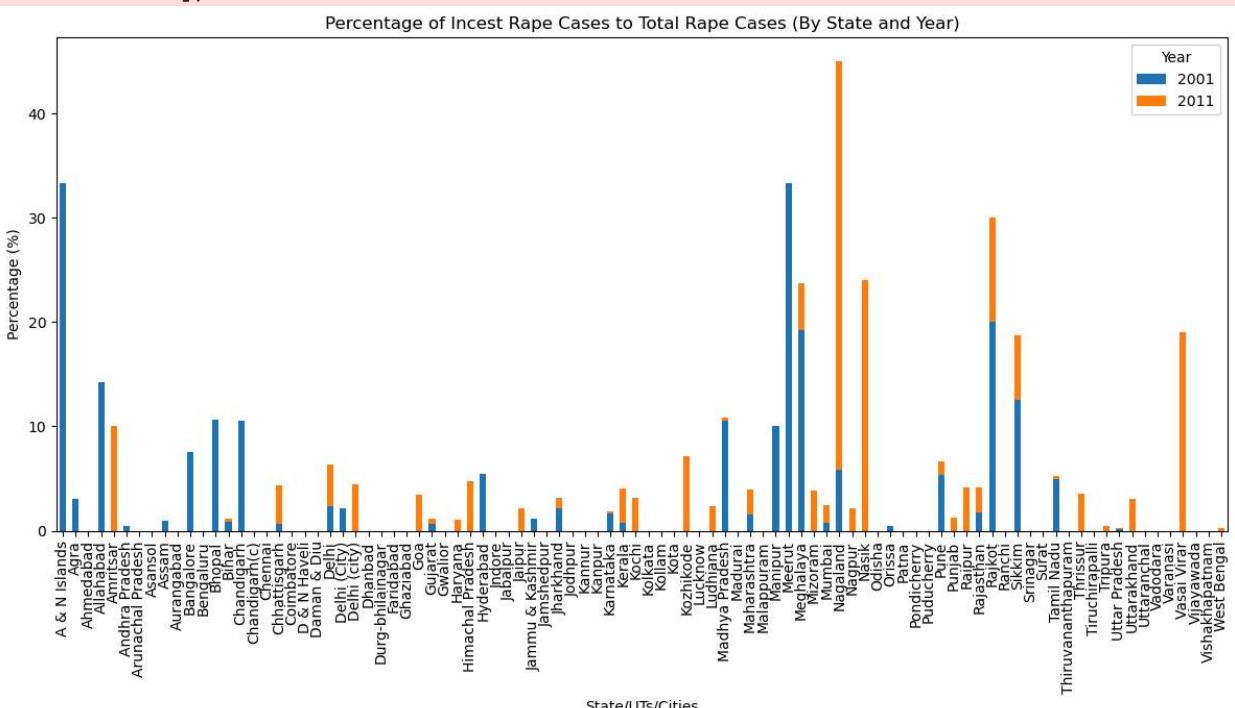
plt.title("Percentage of Incest Rape Cases to Total Rape Cases (By State and Year)")
plt.xlabel("State/UTs/Cities")
plt.ylabel("Percentage (%)")
plt.xticks(rotation=90)
plt.legend(title='Year')
plt.tight_layout()
plt.show()
```

C:\Users\praya\AppData\Local\Temp\ipykernel_27656\891479197.py:4: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
filtered_df['Incest to Total Rape Cases (%)'] = (filtered_df['No. of Victims (Incest Rape Cases) - Total Victims'] / filtered_df['No. of Victims (Total Rape Cases) - Total Victims']) * 100
```



In [121...]

```
# Iterate over each year (column) in pivot_df
for year in pivot_df.columns:
```

```
max_state = pivot_df[year].idxmax()
max_value = pivot_df[year].max()
print(f'{year} → State with maximum percentage of incest rape cases: {max_state},'
      f'2001 → State with maximum percentage of incest rape cases: A & N Islands, Percentage: '
      f'33.33%'
      f'2011 → State with maximum percentage of incest rape cases: Nagaland, Percentage: 39.1 '
      f'3%
```

In []: