

An abstract graphic on the left side of the slide. It features a dark purple L-shaped axis with arrows at the ends. There are five irregular, organic shapes in various colors: a light orange one at the top left, a medium purple one below it, a dark purple one to the left of the center, a small orange one to the right of the center, and a large light purple one at the bottom left.

# Identifying Patterns of **Crime** Against **Women** in India

by Adeline Liem

CSS Summer 2024

# Introduction

# Background

**+26.35%**

increase in reported crimes against women from 2016-2021

**428,278 cases**

of crimes against women in 2021, out of a total 6 mil. reported crimes

# About the Dataset

**34 states** in India examined across **2001-2021**

Records occurrences of crimes against women, including **rape, kidnapping/abduction, dowry deaths, assault on women, domestic violence, and women trafficking**

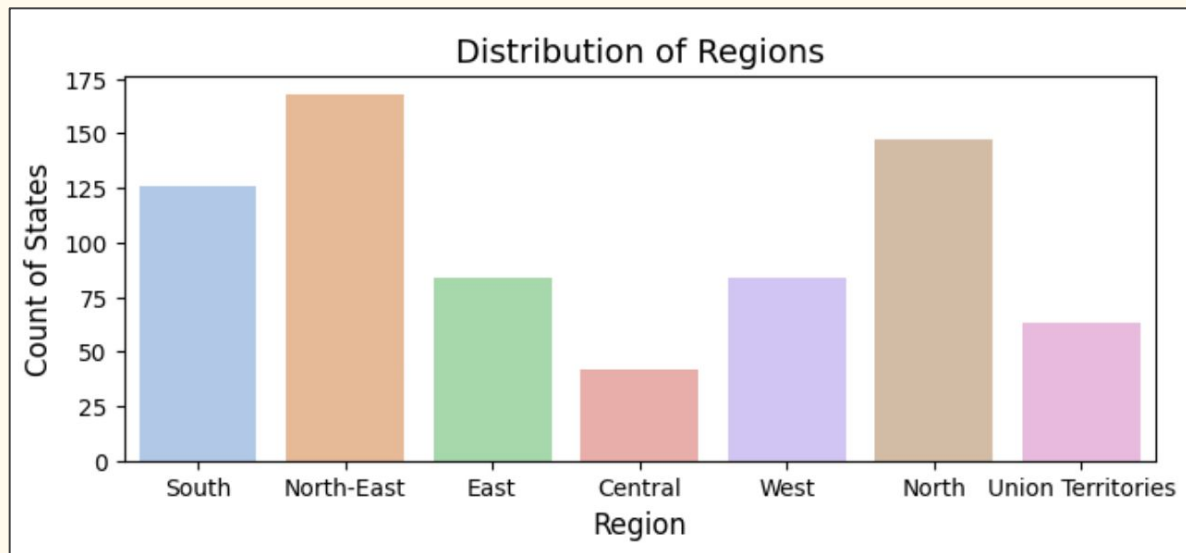


**Which regions exhibit similar patterns of these crimes?**

**Can states be categorized into distinct clusters based on the type and frequency of crimes against women?**

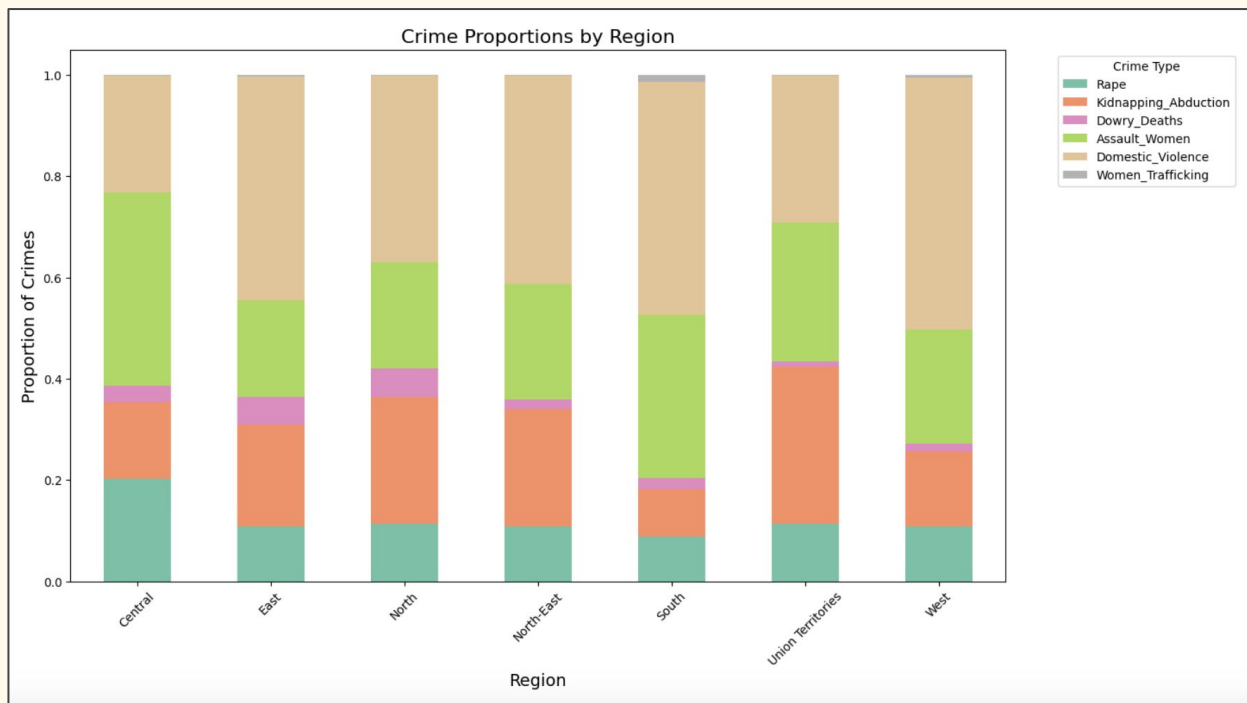
# Exploratory Data Analysis

# Distribution of Regions



- Regions with more states might skew overall statistics
- Could be supplemented with population data

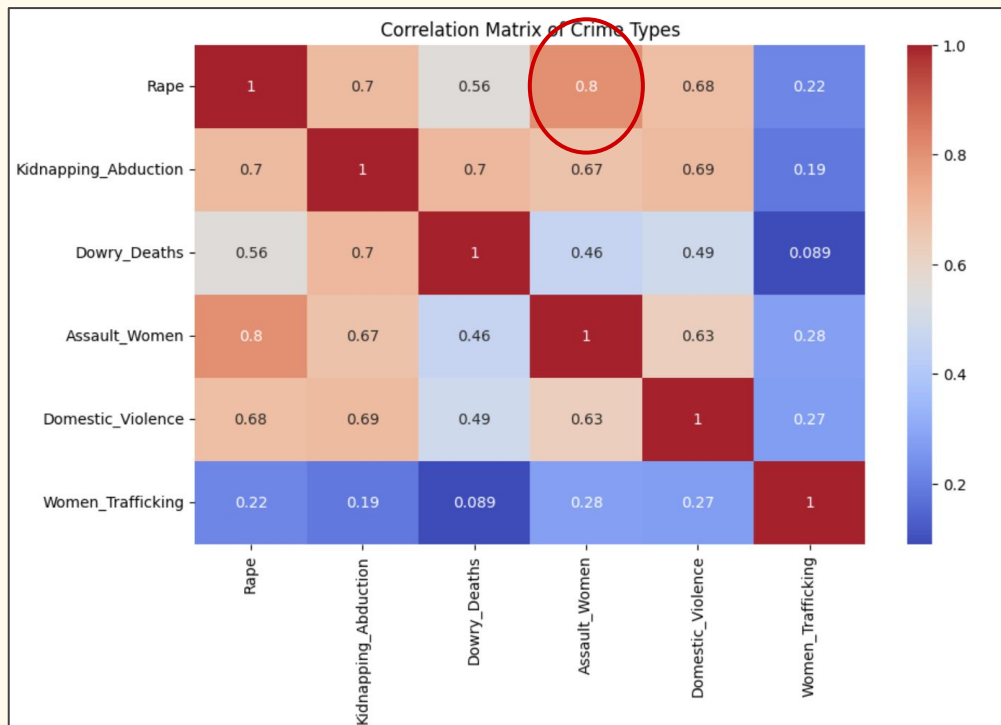
# Crime Proportions by Region



- Proportion of each crime out of the total cases of crimes against women
- Highest tends to be domestic violence, lowest is women trafficking



# Correlation Matrix of Crime Types



- Highest correlation between rape and assault on women
- Generally high correlations between rape and kidnapping/abduction, as well as kidnapping/abduction and dowry deaths ( $r \geq 0.7$ )

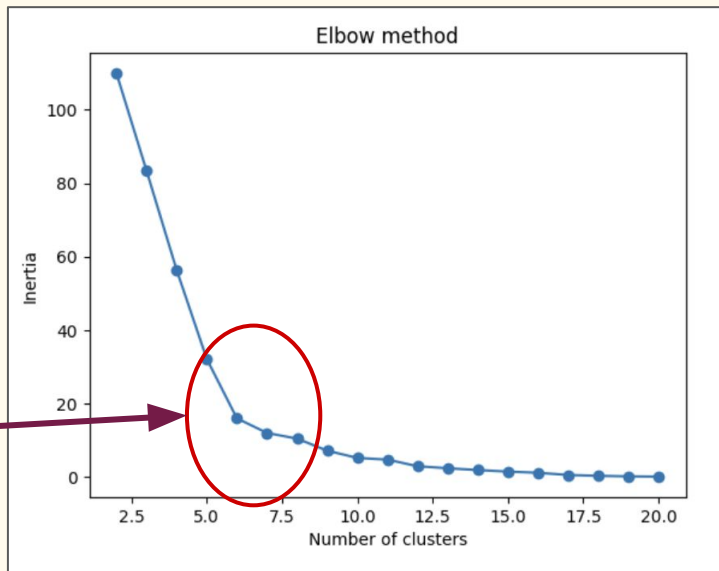
# K-Means Clustering

# Finding K - 2001

The “**elbow method**” is an approach for determining the number of optimal clusters, which involves identifying the smallest value of that significantly reduces within-cluster variation.

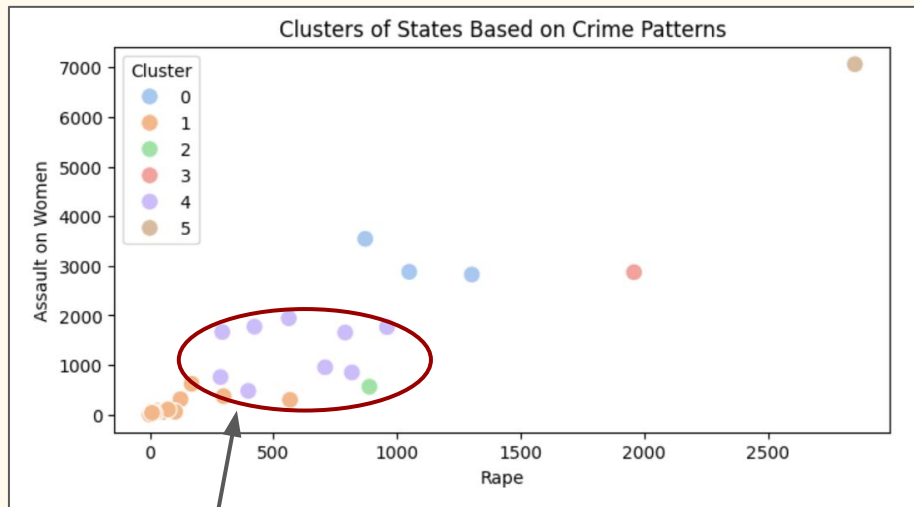
find where the line makes the elbow shape!

$k = 6$

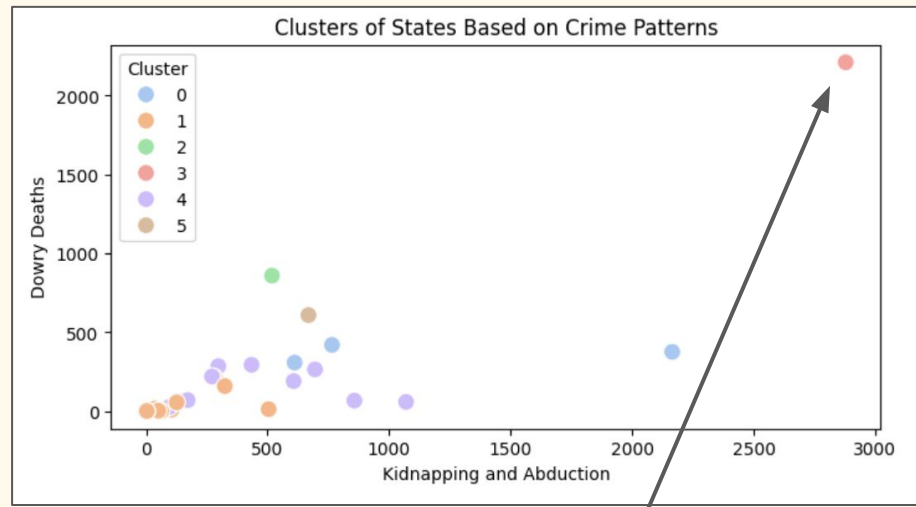


# Scatterplots of Clusters - 2001

Each point represents a state, and the colors denote the different clusters identified by K-Means clustering.



noticeable grouping!



some clusters seem to only have one state

# Crime Occurrence Averages by Cluster - 2001

0.

- 19 states, diff. regions
- Overall high averages across crimes
- Notably high in assault (3081.67) and domestic violence (5804.34)

1.

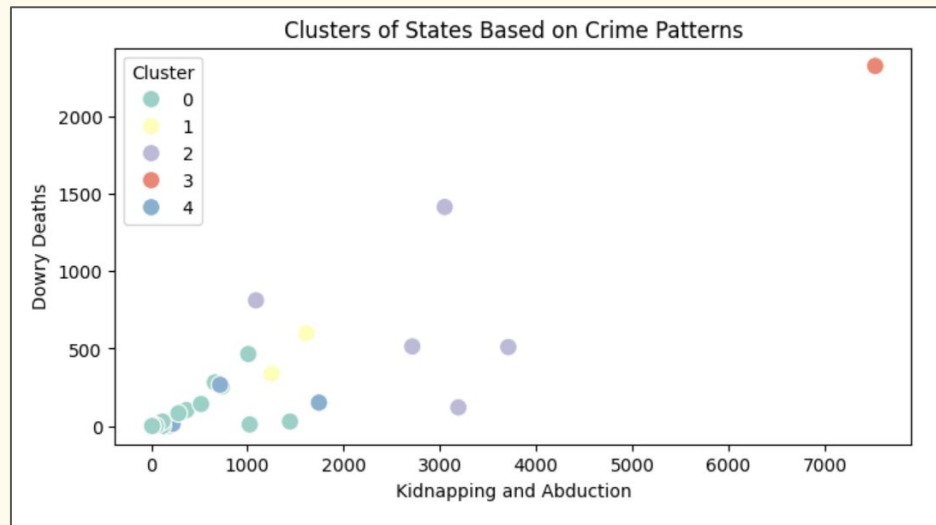
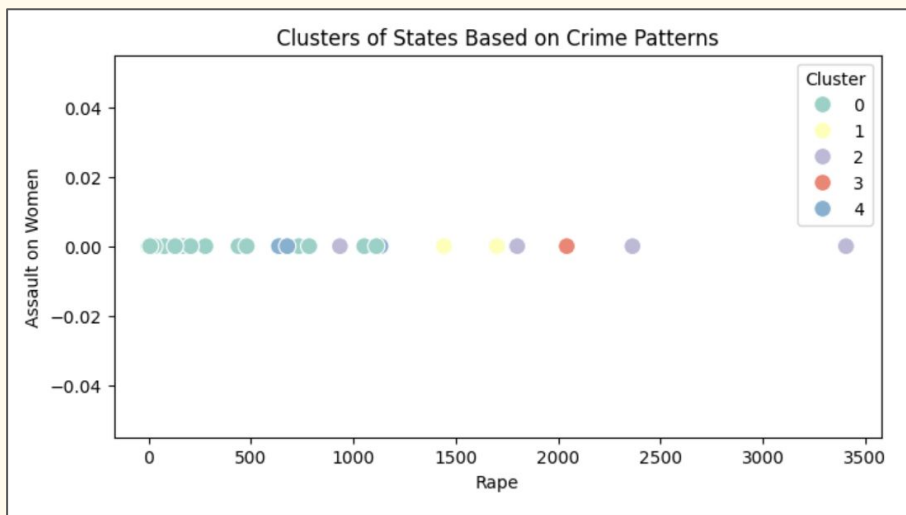
- 3 states from South and West
- Represents states with **lowest occurrences of crimes against women**
- Lowest averages of crimes across all clusters, except for women trafficking (0.26))

3.

- 1 state in the North
- **Highest crime overall**
- Highest averages for kidnapping/abduction (2879), dowry deaths (2211), and domestic violence (7365)

# Scatterplots of Clusters - 2011

Each point represents a state, and the colors denote the different clusters identified by K-Means clustering. (k=5)



# Crime Occurrence Averages by Cluster - 2011

0.

- Lowest occurrences of crimes against women
- High in domestic violence (671.78)

2.

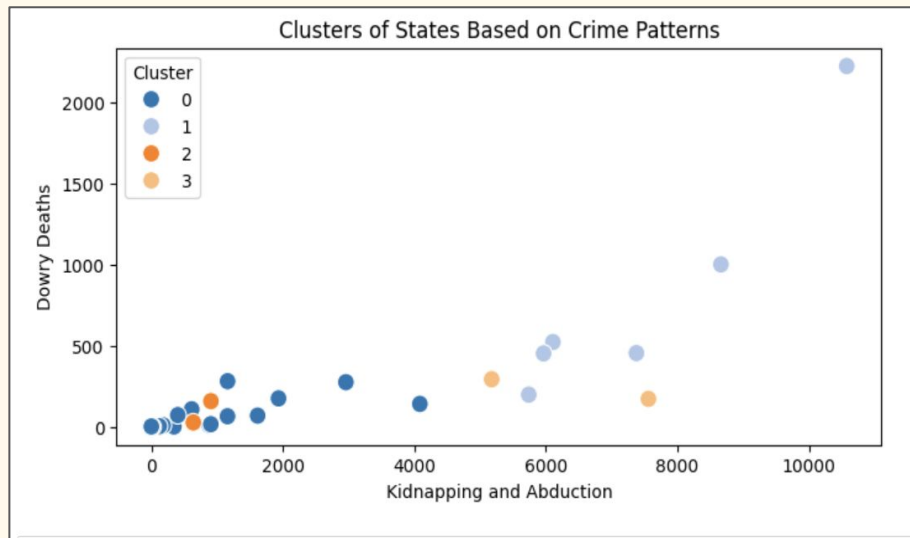
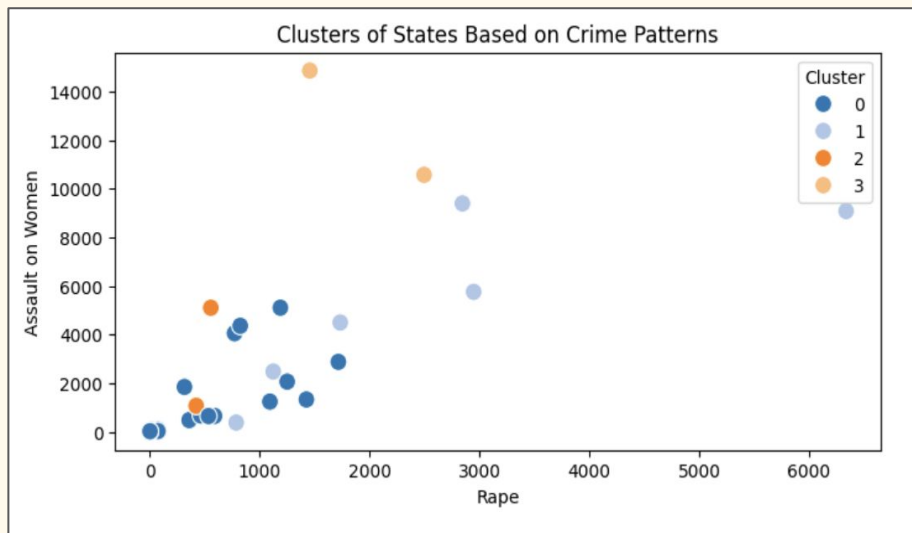
- Moderate to high averages for most crime types
- High in kidnapping/abduction (2750.8), dowry deaths (673.8), and domestic violence (10256)

3.

- Highest averages out of all the clusters for rape (2042), kidnapping/abduction (7525), and dowry deaths (2322)
- Extremely high domestic violence (7121)

# Scatterplots of Clusters - 2021

Each point represents a state, and the colors denote the different clusters identified by K-Means clustering. (k=4)





# Crime Occurrence Averages by Cluster - 2021

0.

- Low-moderate averages of dowry deaths (51.38) and women trafficking (8.46)
- Moderate-high for other crimes

1.

- **Highest averages across multiple crime types** compared to the other clusters
- Especially domestic violence (13037.33) and assault on women (5267.17)

3.

- Highest average for assault on women (12710.5)

# Results

# Can we identify regions with similar crime patterns?

... No, not really. Here's an example from 2001:

**Cluster 0:** 19 states (7 Northeast, 5 North, 1 West, 1 East, 3 Union Territories, 2 South)

**Cluster 1:** 3 states (1 South, 2 West)

**Cluster 2:** 1 state (1 East)

**Cluster 3:** 1 state (1 North)

**Cluster 4:** 1 state (1 Central)

**Cluster 5:** 9 states (3 South, 1 North, 1 Northeast, 2 East, 1 West, 1 Central)

# ... but can we potentially identify states with similar crime patterns?

Yes! For example...

- **Maharashtra and Andrah Pradesh** are in the same cluster type in all 3 years
  - Had notably high counts of domestic violence in 2011 and 2021
- **Uttar Pradesh** had extremely high counts of rape, kidnapping/abduction, and dowry deaths, and domestic violence in years 2001 and 2011
- **Karnataka, Tamil Nadu, and Kerala** all had one of the highest averages in women trafficking in 2011
- **Uttarakhand, Tripura, Kerala, Punjab, Bihar, and Assam** had extremely high levels of domestic violence, assault on women, and dowry deaths in 2021

# Discussion

# Limitations

## 1. Lack of population data

- a. Population data could have added more context into how much crime was occurring per state vs. population per state

## 2. Subjective labels

- a. The averages of the various crime types were labeled as low, moderate, or high, which are subjective and not clearly defined

## 3. Data quality

- a. Different states may have different reporting difficulties, which may impact what crimes are recorded