

Homework 1: USArrests Dataset Analysis

Advanced Data Programming with R

AUTHOR
Anekha Kallath (24205324)

PUBLISHED
June 4, 2025

Introduction

The in-built R dataset `USArrests` is used for this homework. This dataset contains crime statistics for **50 US states** from **1973**.

The dataset contains statistics on:

- **Murder, Assault, and Rape** (arrest rates per 100,000 people)
- **UrbanPop** : percentage of urban population

Murder		Assault		UrbanPop		Rape	
Min.	: 0.800	Min.	: 45.0	Min.	:32.00	Min.	: 7.30
1st Qu.:	4.075	1st Qu.:	109.0	1st Qu.:	54.50	1st Qu.:	15.07
Median	: 7.250	Median	:159.0	Median	:66.00	Median	:20.10
Mean	: 7.788	Mean	:170.8	Mean	:65.54	Mean	:21.23
3rd Qu.:	11.250	3rd Qu.:	249.0	3rd Qu.:	77.75	3rd Qu.:	26.18
Max.	:17.400	Max.	:337.0	Max.	:91.00	Max.	:46.00

Summary:

- **Murder** rates range from **0.8** to **17.4**, with a mean of **7.79**
- **Assault** is the most prevalent crime, with values ranging from **45** to **337**, and a high mean of **170.8**
- **UrbanPop**, the percentage of people living in urban areas, ranges from **32%** to **91%**, with a median of **66%**
- **Rape** arrests range from **7.3** to **46.0**, with a mean of **21.23**

These statistics suggest that **assault** is the dominant form of violent crime in the dataset, while urbanization levels vary considerably across the 50 states.

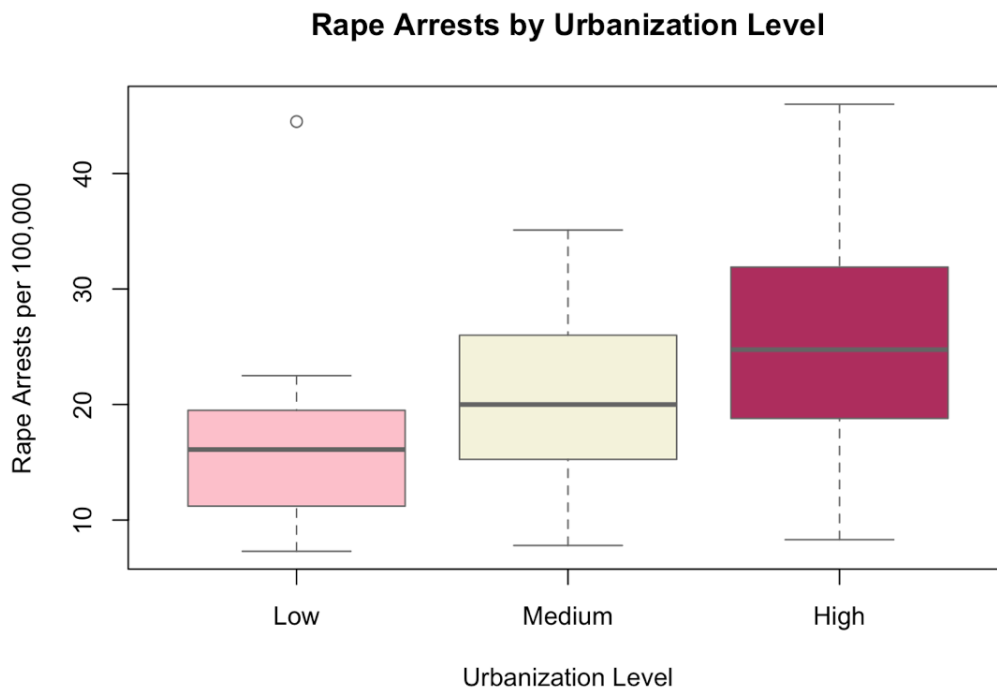
Grouping States by Urban Population

Low	Medium	High
9	27	14

- **Low:** 9 states with low urban population
- **Medium:** 27 states with moderate urban population
- **High:** 14 states with high urban population

Urban Population vs Rape Arrests Plot

The urban population categorised into the three groups previously is plotted against the arrests registered for rape cases.



Conclusion:

The boxplot compares **rape arrest rates** (per 100,000 residents) across three levels of urbanization:

- **Low Urbanization** states tend to have the lowest rape arrest rates with a tighter spread and a lower median.
- **Medium Urbanization** states show a wider range of rape arrest rates, including one or two higher outliers.
- **High Urbanization** states generally have **higher median rape arrest rates** and a broader spread, indicating greater variability and higher extremes.

There is a visible trend where **rape arrest rates increase** with **higher urban population** levels. This pattern may reflect increased reporting,