The Data

The data set contains statistics - in arrests per 100,000 residents – for assault, murder, and rape in each of the 50 US states in 1973. Also given is the percent of the population living in urban areas.

Columns

1. state
2. murder – number of murders arrests per 100,000 residents
3. assault – number of assault arrests per 100,000 residents
4. urbanpop – percentage of the state's population that resides in urban areas
5. rape – number of rape arrests per 100,000 residents

Questions

1. Load the data into your R session by first setting the working directory to the course's folder. Investigate the structure of the dataset
    a. What is the datatype for each data frame column?
    b. How many records are contained in the data?
    c. What is the maximum "urbanpop" value?
2. Let's see if there is a relationship between the urbanpop and assault.
    a. Create a ggplot object by setting aes() with x=urbanpop and y=assault
    b. Add a geom_point() layer to create a scatterplot that shows the relationship between the two variables.
    c. Write down some of your observations.
3. It might make sense to add another layer of categorization to this data. Let us separate the states into 5 regions:
    a. East Coast
    b. West Coast
    c. Mid West
    d. South
    e. Other