

CSCI 3900C Lab #3 (25 points)

Write an R script to accomplish the following tasks. Please put all code in a single script file, and use comments to identify the different sections. Also use comments for your name at the top.

UPLOAD YOUR SCRIPT TO THE LAB3 D2L Dropbox.

1. Load the USArrests data set and create one additional column. The column should be a factor with four levels, where the levels describe the urban population, based on the UrbanPop variable that is already a column in the data frame. Each row should be assigned a level based on the value of UrbanPop and its quartiles, as follows:

Criterion	Level
UrbanPop < Q1	LOW
$Q1 \leq \text{UrbanPop} < \text{Median}$	MODERATE
$\text{Median} \leq \text{UrbanPop} < Q3$	HIGH
$\text{UrbanPop} \geq Q3$	VERY HIGH

2. Use the lattice package to create the following plots.
 - a. Histograms of the number of arrests for assault (per 100,000) grouped by Urban population level (plot should contain 4 histograms on the same scale). Use a heavy red *dashed* vertical line on each histogram to show the median number of assaults.
 - b. Box plots of the number of arrests for assault (per 100,000) grouped by Urban population level. Use a *dotted* blue line on each boxplot to show the mean number of assaults.
 - c. Scatterplots with assault as the predictor (independent) variable and murder as the outcome (dependent) variable, grouped by Urban population level. Use a heavy solid line to show the least squares regression line for each scatterplot.

Suggested approach for step #1:

Add the AssignLevel function shown below to your script; use it with **sapply** to create the new column as a vector. Then convert the vector to a factor, **and designate the appropriate order of levels** (LOW, MODERATE, HIGH, VERY HIGH). Then add the column to the data frame using **cbind**.

```
AssignLevel <- function(p, quartiles)
```

```
{  
  if (p < quartiles[1])  
    rlevel <- "LOW"  
  else if (p < quartiles[2])  
    rlevel <- "MODERATE"  
  else if (p < quartiles[3])  
    rlevel <- "HIGH"  
  else  
    rlevel <- "VERY HIGH"  
  
  return (rlevel)  
}
```

Sample use of this function:

```
> n <- runif(100,1,10)  
> head(n)  
[1] 7.369 7.644 9.514 9.826 8.178 2.517  
> k <- n[20]  
> k  
[1] 6.080374  
> q <- quantile(n, c(.25, .5, .75))  
> q  
      25%      50%      75%  
3.605503 5.795623 7.761620  
> AssignLevel(k, q)  
[1] "HIGH"
```

(continued, next page)

When you have accomplished step 1, the revised data frame should have this structure:

	Murder	Assault	UrbanPop	Rape	UrbanPopRating
Alabama	13.2	236	58	21.2	MODERATE
Alaska	10.0	263	48	44.5	LOW
Arizona	8.1	294	80	31.0	VERY HIGH
Arkansas	8.8	190	50	19.5	LOW
California	9.0	276	91	40.6	VERY HIGH
Colorado	7.9	204	78	38.7	VERY HIGH

(etc.)

Also note that the order of your factor levels will determine the order in which the plots are shown:

