Boston

PRIYA(1948050)

31 January 2020

bdata<-read.csv("C:/Users/PRIYA/Desktop/crime.csv")  
View(bdata)  
str(bdata)

## 'data.frame': 327820 obs. of 17 variables:  
## $ INCIDENT\_NUMBER : Factor w/ 290156 levels "142052550","I010370257-00",..: 290156 290155 290154 290153 290152 290151 290150 290149 290148 290147 ...  
## $ OFFENSE\_CODE : int 2403 3201 2647 413 3122 1402 3803 3301 802 3410 ...  
## $ OFFENSE\_CODE\_GROUP : Factor w/ 67 levels "Aggravated Assault",..: 15 53 47 1 2 64 44 65 62 63 ...  
## $ OFFENSE\_DESCRIPTION: Factor w/ 244 levels "A&B HANDS, FEET, ETC. - MED. ATTENTION REQ.",..: 63 187 222 13 5 231 162 232 21 223 ...  
## $ DISTRICT : Factor w/ 13 levels "","A1","A15",..: 12 9 5 2 4 7 1 5 12 10 ...  
## $ REPORTING\_AREA : int 495 795 329 92 36 351 NA 603 543 621 ...  
## $ SHOOTING : Factor w/ 2 levels "","Y": 1 1 1 1 1 1 1 1 1 1 ...  
## $ OCCURRED\_ON\_DATE : Factor w/ 239364 levels "2015-06-15 00:00:00",..: 239362 232564 239354 239361 239364 239180 239363 239357 239355 239361 ...  
## $ YEAR : int 2018 2018 2018 2018 2018 2018 2018 2018 2018 2018 ...  
## $ MONTH : int 10 8 10 10 10 10 10 10 10 10 ...  
## $ DAY\_OF\_WEEK : Factor w/ 7 levels "Friday","Monday",..: 7 5 7 7 7 6 7 7 7 7 ...  
## $ HOUR : int 20 20 19 20 20 20 20 19 19 20 ...  
## $ UCR\_PART : Factor w/ 5 levels "","Other","Part One",..: 5 4 5 3 4 5 4 4 5 4 ...  
## $ STREET : Factor w/ 4685 levels ""," ALBANY ST ",..: 242 145 1275 730 3410 1317 1 4213 307 1025 ...  
## $ Lat : num 42.3 42.4 42.3 42.4 42.4 ...  
## $ Long : num -71.1 -71.1 -71.1 -71.1 -71 ...  
## $ Location : Factor w/ 18255 levels "(-1.00000000, -1.00000000)",..: 906 14711 7033 15950 17227 5997 9166 11261 606 13936 ...

dim(bdata)

## [1] 327820 17

summary(bdata)

## INCIDENT\_NUMBER OFFENSE\_CODE   
## I162030584: 13 Min. : 111   
## I152080623: 11 1st Qu.:1001   
## I172013170: 10 Median :2907   
## I172096394: 10 Mean :2318   
## I182065208: 10 3rd Qu.:3201   
## I162001871: 9 Max. :3831   
## (Other) :327757   
## OFFENSE\_CODE\_GROUP  
## Motor Vehicle Accident Response: 38134   
## Larceny : 26670   
## Medical Assistance : 24226   
## Investigate Person : 19176   
## Other : 18612   
## Drug Violation : 17037   
## (Other) :183965   
## OFFENSE\_DESCRIPTION DISTRICT   
## SICK/INJURED/MEDICAL - PERSON : 19360 B2 :51288   
## INVESTIGATE PERSON : 19180 C11 :43817   
## M/V - LEAVING SCENE - PROPERTY DAMAGE: 16730 D4 :43338   
## VANDALISM : 15542 A1 :36735   
## ASSAULT SIMPLE - BATTERY : 15191 B3 :36400   
## VERBAL DISPUTE : 13478 C6 :24190   
## (Other) :228339 (Other):92052   
## REPORTING\_AREA SHOOTING OCCURRED\_ON\_DATE YEAR   
## Min. : 0.0 :326765 2016-08-01 00:00:00: 29 Min. :2015   
## 1st Qu.:177.0 Y: 1055 2017-06-01 00:00:00: 29 1st Qu.:2016   
## Median :343.0 2015-07-01 00:00:00: 27 Median :2017   
## Mean :383.2 2015-06-18 05:00:00: 22 Mean :2017   
## 3rd Qu.:544.0 2017-08-01 00:00:00: 22 3rd Qu.:2017   
## Max. :962.0 2017-01-01 00:00:00: 21 Max. :2018   
## NA's :20920 (Other) :327670   
## MONTH DAY\_OF\_WEEK HOUR UCR\_PART   
## Min. : 1.000 Friday :49758 Min. : 0.00 : 93   
## 1st Qu.: 4.000 Monday :46970 1st Qu.: 9.00 Other : 1285   
## Median : 7.000 Saturday :45969 Median :14.00 Part One : 63231   
## Mean : 6.672 Sunday :41374 Mean :13.11 Part Three:162928   
## 3rd Qu.: 9.000 Thursday :47872 3rd Qu.:18.00 Part Two :100283   
## Max. :12.000 Tuesday :47726 Max. :23.00   
## Wednesday:48151   
## STREET Lat Long   
## WASHINGTON ST : 14590 Min. :-1.00 Min. :-71.18   
## : 10977 1st Qu.:42.30 1st Qu.:-71.10   
## BLUE HILL AVE : 8002 Median :42.33 Median :-71.08   
## BOYLSTON ST : 7425 Mean :42.21 Mean :-70.91   
## DORCHESTER AVE: 5297 3rd Qu.:42.35 3rd Qu.:-71.06   
## TREMONT ST : 4925 Max. :42.40 Max. : -1.00   
## (Other) :276604 NA's :20632 NA's :20632   
## Location   
## (0.00000000, 0.00000000) : 20632   
## (42.34862382, -71.08277637): 1276   
## (42.36183857, -71.05976489): 1248   
## (42.28482577, -71.09137369): 1137   
## (42.32866284, -71.08563401): 1075   
## (42.25621592, -71.12401947): 916   
## (Other) :301536

colSums(is.na(bdata))

## INCIDENT\_NUMBER OFFENSE\_CODE OFFENSE\_CODE\_GROUP   
## 0 0 0   
## OFFENSE\_DESCRIPTION DISTRICT REPORTING\_AREA   
## 0 0 20920   
## SHOOTING OCCURRED\_ON\_DATE YEAR   
## 0 0 0   
## MONTH DAY\_OF\_WEEK HOUR   
## 0 0 0   
## UCR\_PART STREET Lat   
## 0 0 20632   
## Long Location   
## 20632 0

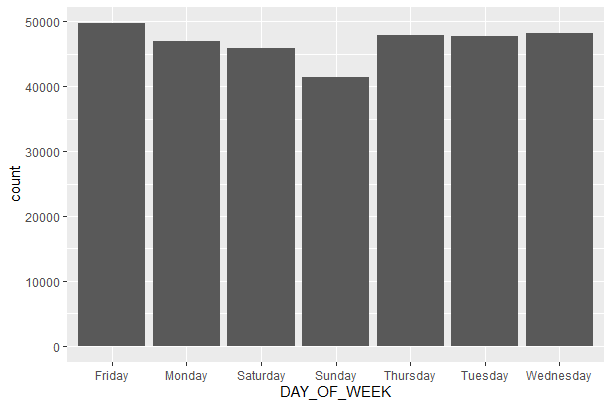
bdata[][is.na(bdata[])] <- 0  
colSums(is.na(bdata))

## INCIDENT\_NUMBER OFFENSE\_CODE OFFENSE\_CODE\_GROUP   
## 0 0 0   
## OFFENSE\_DESCRIPTION DISTRICT REPORTING\_AREA   
## 0 0 0   
## SHOOTING OCCURRED\_ON\_DATE YEAR   
## 0 0 0   
## MONTH DAY\_OF\_WEEK HOUR   
## 0 0 0   
## UCR\_PART STREET Lat   
## 0 0 0   
## Long Location   
## 0 0

library(ggplot2)

ggplot(bdata, aes(x=DAY\_OF\_WEEK)) +

geom\_bar()

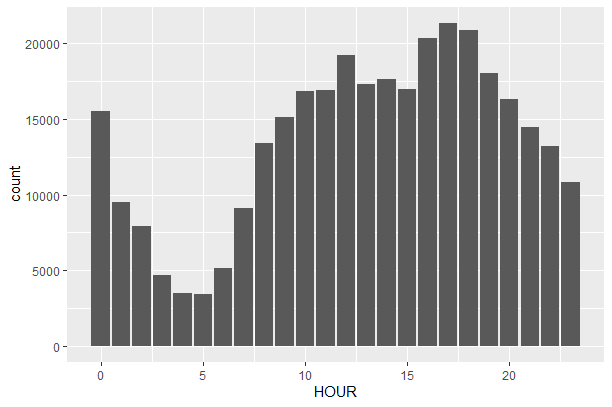


CONCLUSION:

Boston is the capital and most populous city of the Commonwealth of Massachusetts in the United States, and the 21st most populous city in the United States. The percentage of college-educated young adults is larger in the Boston area than anywhere else in America. When we plot the graph for crime happening in each day for the Boston crime dataset, it is seen that the crime rate is less during Sunday. As Sunday is the weekend for most of the employees and students, the presence of people around will be more during weekend. Therefore it is a difficult task for the criminals to commit a crime.

ggplot(bdata, aes(x=HOUR)) +

geom\_bar()



CONCLUSION:

The crime rate is more during the evening time and is gradually decreasing by time when we look at the plot plotted with respect to Hours.