An Analytic Detective

Hasan Ali ÖZKAN

11/9/2021

```
mvt <- read.csv("mvtWeek1.csv",sep = ",",header = T)
```

How many rows of data (observations) are in this dataset?

```
print(paste("There are ",nrow(mvt) ,"observations."))
```

```
## [1] "There are  191641 observations."
```

How many variables are in this dataset?

```
print(paste("There are ",ncol(mvt) ,"columns."))
```

```
## [1] "There are  11 columns."
```

Using the “max” function, what is the maximum value of the variable “ID”?

```
print(paste("The maximum value of the 'ID' is ",max(mvt$ID)))
```

```
## [1] "The maximum value of the 'ID' is  9181151"
```

What is the minimum value of the variable “Beat”?

```
print(paste("The minimum value of the 'Beat' is ",min(mvt$Beat)))
```

```
## [1] "The minimum value of the 'Beat' is  111"
```

How many observations have value TRUE in the Arrest variable (this is the number of crimes for which an arrest was made)?

```
print(paste(nrow(subset(mvt,mvt$Arrest==TRUE))," have the value of TRUE in the arrest va
riable."))
```

```
## [1] "15536  have the value of TRUE in the arrest variable."
```

How many observations have a LocationDescription value of ALLEY?

```
print(paste(nrow(subset(mvt,mvt$LocationDescription=="ALLEY"))," have the value of ALLEY
in the LocationDescription variable."))
```

```
## [1] "2308  have the value of ALLEY in the LocationDescription variable."
```

In what format are the entries in the variable Date?

```
print(paste(mvt$Date[1], "as you see the date format is Month/Day/Year Hour:Minute"))
```

```
## [1] "12/31/12 23:15 as you see the date format is Month/Day/Year Hour:Minute"
```

What is the month and year of the median date in our dataset?

```
DateConvert = as.Date(strptime(mvt$Date, "%m/%d/%y %H:%M"))
median(DateConvert)
```

```
## [1] "2006-05-21"
```

As you see the answer is ***May 2006***.

In which month did the fewest motor vehicle thefts occur?

```
mvt$Month = months(DateConvert)
mvt$Weekday = weekdays(DateConvert)
mvt$Date = DateConvert
table(mvt$Month)
```

```
##
##     April    August  December  February   January      July      June     March
##     15280     16572     16426     13511     16047     16801     16002     15758
##       May  November   October September
##     16035     16063     17086     16060
```

As you see the answer is ***February***.

On which weekday did the most motor vehicle thefts occur?

```
table(mvt$Weekday)
```

```
##
##    Friday    Monday  Saturday    Sunday  Thursday   Tuesday Wednesday
##     29284     27397     27118     26316     27319     26791     27416
```

```
max(table(mvt$Weekday))
```

```
## [1] 29284
```

As you see the answer is ***Friday***.

Which month has the largest number of motor vehicle thefts for which an arrest was made?

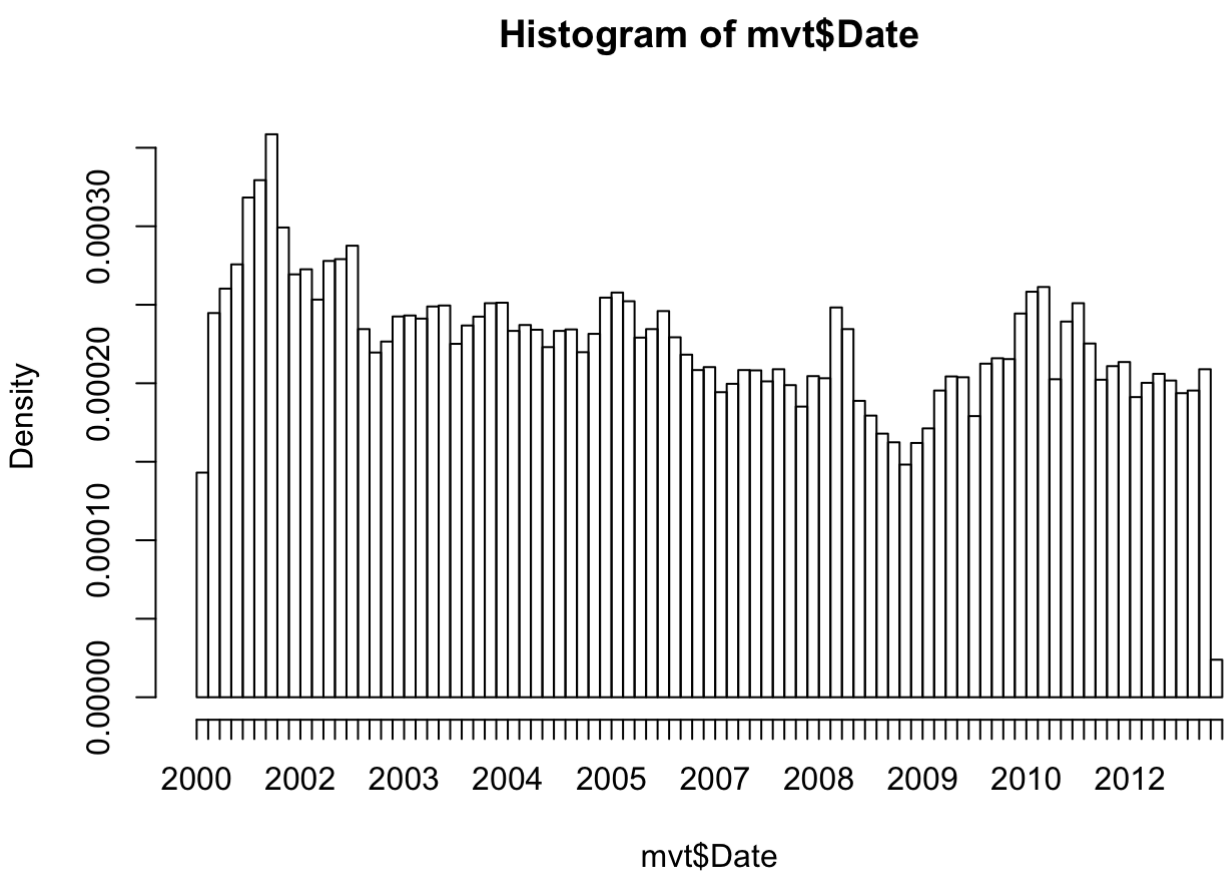
```
table(mvt$Month, mvt$Arrest)
```

```
##
##              FALSE  TRUE
##   April      14028  1252
##   August     15243  1329
##   December   15029  1397
##   February   12273  1238
##   January    14612  1435
##   July       15477  1324
##   June       14772  1230
##   March      14460  1298
##   May        14848  1187
##   November   14807  1256
##   October    15744  1342
##   September 14812  1248
```

As tou see the answer is ***January***.

Looking at the histogram, answer the following questions.

```
hist(mvt$Date,breaks = 100)
```



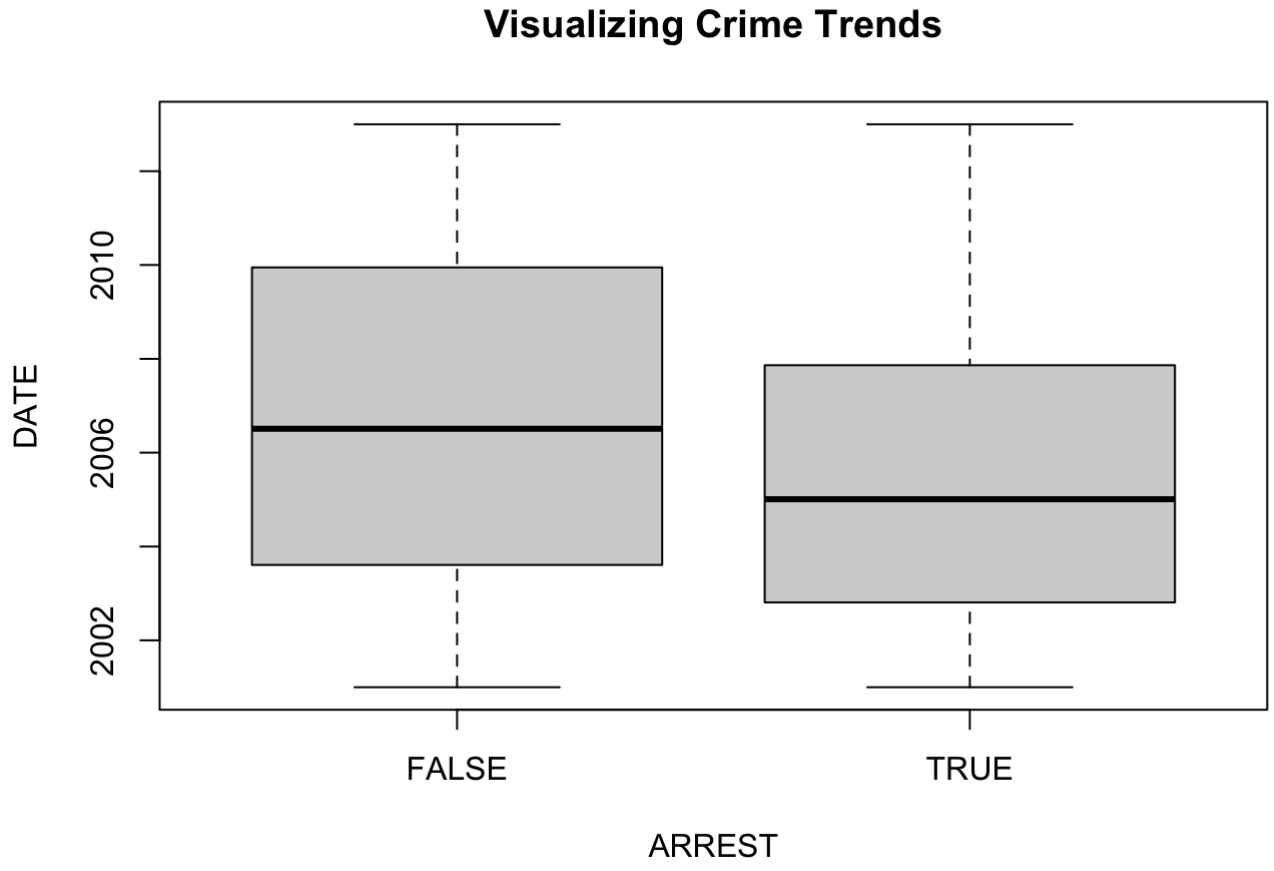
In general, does it look like crime increases or decreases from 2002 - 2012?
Increases
Decreases

In general, does it look like crime increases or decreases from 2005 - 2008?
Increases
Decreases

In general, does it look like crime increases or decreases from 2009 - 2011?
Increases
Decreases

Does it look like there were more crimes for which arrests were made in the first half of the time period or the second half of the time period? (Note that the time period is from 2001 to 2012, so the middle of the time period is the beginning of 2007.)

```
boxplot(mvt$Date ~ mvt$Arrest,ylab = "DATE",xlab = "ARREST",main="Visualizing Crime Tren
ds")
```



As you see above the answer is ***first half***.

For what proportion of motor vehicle thefts in 2001 was an arrest made?

```
table(mvt$Year,mvt$Arrest)
```

```
##
##        FALSE  TRUE
##   2001 18517  2152
##   2002 16638  2115
##   2003 14859  1798
##   2004 15169  1693
##   2005 14956  1528
##   2006 14796  1302
##   2007 13068  1212
##   2008 13425  1020
##   2009 11327   840
##   2010 14796   701
##   2011 15012   625
##   2012 13542   550
```

```
print(paste("The answer is ",2152/(2152+18517)))
```

```
## [1] "The answer is  0.10411727708162"
```

For what proportion of motor vehicle thefts in 2007 was an arrest made?

```
table(mvt$Year,mvt$Arrest)
```

```
##
##        FALSE  TRUE
##   2001 18517  2152
##   2002 16638  2115
##   2003 14859  1798
##   2004 15169  1693
##   2005 14956  1528
##   2006 14796  1302
##   2007 13068  1212
##   2008 13425  1020
##   2009 11327   840
##   2010 14796   701
##   2011 15012   625
##   2012 13542   550
```

```
print(paste("The answer is ",1212/(1212+13068)))
```

```
## [1] "The answer is  0.0848739495798319"
```

For what proportion of motor vehicle thefts in 2012 was an arrest made?

```
table(mvt$Year,mvt$Arrest)
```

```
##
##        FALSE  TRUE
##   2001 18517  2152
##   2002 16638  2115
##   2003 14859  1798
##   2004 15169  1693
##   2005 14956  1528
##   2006 14796  1302
##   2007 13068  1212
##   2008 13425  1020
##   2009 11327   840
##   2010 14796   701
##   2011 15012   625
##   2012 13542   550
```

```
print(paste("The answer is ",550/(550+13542)))
```

```
## [1] "The answer is  0.0390292364462106"
```

Which locations are the top five locations for motor vehicle thefts, excluding the "Other" category? You should select 5 of the following options.

```
sort(table(mvt$LocationDescription),decreasing = T)[1:6]
```

```
##
##                         STREET PARKING LOT/GARAGE(NON.RESID.)
##                         156564                          14852
##                          OTHER                          ALLEY
##                           4573                           2308
##                    GAS STATION         DRIVEWAY - RESIDENTIAL
##                           2111                           1675
```

How many observations are in Top5?

```
the_top_five <-  subset(mvt, LocationDescription=="STREET" | LocationDescription=="PARKI
NG LOT/GARAGE(NON.RESID.)" | LocationDescription=="ALLEY" | LocationDescription=="GAS ST
ATION" | LocationDescription=="DRIVEWAY - RESIDENTIAL")
print(paste("There are ",nrow(the_top_five) ,"observations."))
```

```
## [1] "There are  177510 observations."
```

One of the locations has a much higher arrest rate than the other locations. Which is it? Please enter the text in exactly the same way as how it looks in the answer options for Problem 4.1.

```
table(the_top_five$LocationDescription, the_top_five$Arrest)
```

```
##
##                                 FALSE   TRUE
##   ALLEY                          2059    249
##   DRIVEWAY - RESIDENTIAL         1543    132
##   GAS STATION                    1672    439
##   PARKING LOT/GARAGE(NON.RESID.)  13249   1603
##   STREET                        144969  11595
```

As you see above it is about ***20%***

On which day of the week do the most motor vehicle thefts at gas stations happen?

```
table(the_top_five$Weekday,the_top_five$LocationDescription)
```

```
##
##             ALLEY DRIVEWAY - RESIDENTIAL GAS STATION
##   Friday      385                    257         332
##   Monday      320                    255         280
##   Saturday    341                    202         338
##   Sunday      307                    221         336
##   Thursday    315                    263         282
##   Tuesday     323                    243         270
##   Wednesday   317                    234         273
##
##             PARKING LOT/GARAGE(NON.RESID.) STREET
##   Friday                             2331  23773
##   Monday                             2128  22305
##   Saturday                           2199  22175
##   Sunday                             1936  21756
##   Thursday                           2082  22296
##   Tuesday                            2073  21888
##   Wednesday                          2103  22371
```

As you see above the answer is ***Saturday***

On which day of the week do the fewest motor vehicle thefts in residential driveways happen?

```
table(the_top_five$Weekday,the_top_five$LocationDescription)
```

```
##
##             ALLEY DRIVEWAY - RESIDENTIAL GAS STATION
##   Friday      385                     257         332
##   Monday      320                     255         280
##   Saturday    341                     202         338
##   Sunday      307                     221         336
##   Thursday    315                     263         282
##   Tuesday     323                     243         270
##   Wednesday   317                     234         273
##
##             PARKING LOT/GARAGE(NON.RESID.) STREET
##   Friday                             2331  23773
##   Monday                             2128  22305
##   Saturday                           2199  22175
##   Sunday                             1936  21756
##   Thursday                           2082  22296
##   Tuesday                            2073  21888
##   Wednesday                          2103  22371
```

As you see above the answer is ***Saturday***