Homework Module 3

word\_document: default pdf\_document: default —

**Author:** Andres Felipe Alba Hernández  
**Department:** Electrical Engineering  
**Date:** September 19, 2018  
**Course:** ISYE670 Data Science for Engineers  
**Professor:** Dr. Christine Nguyen  
**Northern Illinois University**

#Cleaning the enviroment variables  
rm(list=ls())

Chicago\_crime\_data=read.csv(file = "Crimes\_Chicago\_2015.csv",header = TRUE)#Reading data from file and setting that in a dataframe

#Data exploration  
#summary(Chicago\_crime\_data)  
#structure(Chicago\_crime\_data)  
#names(Chicago\_crime\_data)  
#str(Chicago\_crime\_data)

***Solution*** The above information may be use for solving the following questions

Question 1:  
(2 points) What are the variables representing? The variables represent characteristics of crime cases handle by the Chicago police department.

Question 2:  
(2 points) Which variables are factor variables?  
Case Number, Date, Block, IUCR, Primary Type, Description, Location Description, Arrest, Domestic, FBI code, Update on, and Location. ##

Question 3: NA values

1. (3 points) How many total values in the dataset are stored as NA?  
   23810 values in the data set are not available.

sum(is.na(Chicago\_crime\_data))

## [1] 23810

1. (5 points) For each variable, how many total values are stored as NA?

The output below shows the number of Non Available per columns (variable). As may be observed, the majority of the missing data is in X,Y cordinates, and latitude and longitude. Most probably this missing values were not available at the moment of the report. Example: Somebody report an incident through the phone, not unit available to reach the point and it was a minor incident.

apply(is.na(Chicago\_crime\_data),2,sum) #apply(dataset,dimension[2=column],function)

## ID Case.Number Date   
## 0 0 0   
## Block IUCR Primary.Type   
## 0 0 0   
## Description Location.Description Arrest   
## 0 0 0   
## Domestic Beat District   
## 0 0 0   
## Ward Community.Area FBI.Code   
## 2 0 0   
## X.Coordinate Y.Coordinate Year   
## 5952 5952 0   
## Updated.On Latitude Longitude   
## 0 5952 5952   
## Location   
## 0

Question 4: Factor Variables  
a) (2 points) What are the different factor levels for the primary description?

According to the data 24614 are Domestic Battery Simple crimes.

#unique(Chicago\_crime\_data$Primary.Type)  
#unique(Chicago\_crime\_data$Description)  
sum(Chicago\_crime\_data$Description=="DOMESTIC BATTERY SIMPLE")

## [1] 24614

1. (2 points) How many incidents of arson resulted in an arrest?

453 incidents in Arson resulted in an arrest.

sum(Chicago\_crime\_data$Primary.Type=="ARSON"&Chicago\_crime\_data$Arrest!=0)

## [1] 453

1. (2 points) How many incidents of arson were in an abandoned building?

According to the search below 7 incidents where located an abandoned building in ARSON.

#unique(Chicago\_crime\_data$Location.Description)  
sum(Chicago\_crime\_data$Primary.Type=="ARSON"&Chicago\_crime\_data$Location.Description=="ABANDONED BUILDING")

## [1] 7

1. (1 point) How many incidents in the dataset are domestic and resulted in an arrest?

Zero domestic incidents result in arrest.

sum(Chicago\_crime\_data$Description=="DOMESTIC BATTERY SIMPLE"&Chicago\_crime\_data$Arrest!=0)

## [1] 24614

sum(Chicago\_crime\_data$Description=="DOMESTIC BATTERY SIMPLE"&Chicago\_crime\_data$Arrest==0)

## [1] 0

sum(Chicago\_crime\_data$Description=="DOMESTIC BATTERY SIMPLE")

## [1] 24614

1. (1 point) How many non-domestic incidents resulted in an arrest?

239227 non Domestic incidents result in arrest.

sum(Chicago\_crime\_data$Description!="DOMESTIC BATTERY SIMPLE"&Chicago\_crime\_data$Arrest!=0)

## [1] 239227