Loading, Saving, and Describing Data

Mohammed Abdul Bari

2022-06-11

# Description of the data

This dataset contains data about police misconduct and how such misconduct Costs atlanta city millions every year in settlements and in the aspect of ensuring accountability.The data was collected, stored and categorized base on human judgement calls. Amount of money paid due to police misconduct constitutes of guesses for may be the range provided.This data contain descriptions of the types of misconduct and depicts the measure of police misconduct. The data was stored in a csv file whereby values are separated by comma.However,their width is not fixed unlike flat file.

#Reading the data into R

read\_data <- read.csv(url("https://raw.githubusercontent.com/fivethirtyeight/police-settlements/main/atlanta\_ga/final/atlanta\_edited.csv"))  
  
#I have used read.csv() function which is a base R function. This function creates a regular data frame. In addition, it takes three parameters: file location, header, separator  
# file location shows where the csv file if found.  
# header is a boolean which tells if the data contains header row,so the first row is the name of each column   
# separator (sep) it shows the used field separator character   
# declaring a dataframe

#Clean the data

## [1] "claim\_number" "plaintiff\_name" "summary\_allegations"  
## [4] "amount\_awarded" "incident\_date" "closed\_date"   
## [7] "claim\_or\_lawsuit" "city" "state"   
## [10] "denied\_date" "calendar\_year" "incident\_year"   
## [13] "matter\_name" "court" "docket\_number"   
## [16] "filed\_date" "filed\_year" "location"   
## [19] "other\_expenses" "collection" "total\_incurred"   
## [22] "plaintiff\_attorney" "case\_outcome" "amount\_demanded"

#Characteristics of the data

#number of rows  
nrow(read\_data)

## [1] 65

#number of columns  
ncol(read\_data)

## [1] 24

#remove empty columns from the data frame  
empty\_columns <- sapply(read\_data, function(x) all(is.na(x) | x == ""))  
data <- read\_data[, !empty\_columns]  
#columns with data  
ncol(data)

## [1] 17

#remove empty rows from data frame  
data <- data[!apply(data == "", 1, all),]  
#rows with data  
nrow(data)

## [1] 65

col\_name<- colnames(data)  
col\_description <-c(" claim unique idententification code","Name of plaintiff or claimant","Description of allegations"," Amount awarded to claimant in the settlement","Date on which incident took place","claim closing date","legal demands for compensation or legal actions decided in court"," City name","State abbreviation","date denied","calendar year","incident year","matter name", "Court in which the settlement was reached, when available","Case docket number, when available","Case status as of the date the data was collected, e.g. whether a case was settled, went to jury, or is still pending","amount demanded by plaintiff")  
  
data <- data.frame(col\_name, col\_description)  
data

## col\_name  
## 1 claim\_number  
## 2 plaintiff\_name  
## 3 summary\_allegations  
## 4 amount\_awarded  
## 5 incident\_date  
## 6 closed\_date  
## 7 claim\_or\_lawsuit  
## 8 city  
## 9 state  
## 10 denied\_date  
## 11 calendar\_year  
## 12 incident\_year  
## 13 matter\_name  
## 14 court  
## 15 docket\_number  
## 16 case\_outcome  
## 17 amount\_demanded  
## col\_description  
## 1 claim unique idententification code  
## 2 Name of plaintiff or claimant  
## 3 Description of allegations  
## 4 Amount awarded to claimant in the settlement  
## 5 Date on which incident took place  
## 6 claim closing date  
## 7 legal demands for compensation or legal actions decided in court  
## 8 City name  
## 9 State abbreviation  
## 10 date denied  
## 11 calendar year  
## 12 incident year  
## 13 matter name  
## 14 Court in which the settlement was reached, when available  
## 15 Case docket number, when available  
## 16 Case status as of the date the data was collected, e.g. whether a case was settled, went to jury, or is still pending  
## 17 amount demanded by plaintiff

library(knitr)  
kable(data[1:7,],caption="plantiff information")

plantiff information

| col\_name | col\_description |
| --- | --- |
| claim\_number | claim unique idententification code |
| plaintiff\_name | Name of plaintiff or claimant |
| summary\_allegations | Description of allegations |
| amount\_awarded | Amount awarded to claimant in the settlement |
| incident\_date | Date on which incident took place |
| closed\_date | claim closing date |
| claim\_or\_lawsuit | legal demands for compensation or legal actions decided in court |

# Summary statistics

#number of missing values   
 count=0  
 for(i in 1:nrow(read\_data)){  
 for(j in 1:ncol(read\_data)){  
 if(isTRUE(read\_data[i,j]=="")){  
 count = count + 1  
 }  
 }  
 }  
 count

## [1] 344