# Vishesh Gupta. Applicant for Master's in Applied Data Science.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.5     v dplyr   1.0.3
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(readr)
library(dplyr)
#install.packages("stringr")
library("stringr")
urb_chm <- read_csv("https://uofi.box.com/shared/static/b6u4it6rjdjva4ouk4fps7h7ias0u4eu.csv", col_types
#View(urb_chm)
```

```
urb_chm$Latitude = latitude1 = as.numeric(str_extract(urb_chm$`Mappable Address`, "4\\d+\\.\\d+"))
urb_chm$Longitude = longitude1 = as.numeric(str_extract(urb_chm$`Mappable Address`, "\\-\\d+\\.+\\d+"))
mapad = str_replace(urb_chm$`Mappable Address`,"\\(\\-\\d+\\.+\\d+",paste0("(",latitude1))
urb_chm$`Mappable Address` = str_replace(mapad,"4\\d+\\.\\d+\\)", paste0(longitude1,")"))
urb_chm = subset(urb_chm, select = -c(Latitude,Longitude) )
head(urb_chm, n=5)
```

```
## # A tibble: 5 x 10
##   `Property Addre~ `Parcel Number` `Inspection Dat~ `Inspection Yea~ Grade
##   <chr>                     <dbl> <date>                       <dbl> <chr>
## 1 2508 A Anderson~    932121357004 2017-03-27                    2017 Clas~
## 2 1303 Bradley Av~    912107203006 2011-08-12                    2011 Clas~
## 3 505 West Califo~    922117177004 2008-06-02                    2008 Clas~
## 4 904 North Broad~    912108257067 2008-02-04                    2008 Clas~
## 5 107 East Mumfor~    932120254020 2016-09-02                    2016 Clas~
## # ... with 5 more variables: `License Status` <chr>, `Expiration Date` <date>,
## #   `Mappable Address` <chr>, `Zip Codes` <dbl>,
## #   `:@computed_region_3h3r_2q6z` <dbl>
```

```
lat1 <- as.numeric(str_extract(urb_chm$`Mappable Address`, "4\\d+\\.\\d+"))
lon1 <- as.numeric(str_extract(urb_chm$`Mappable Address`, "\\-\\d+\\.+\\d+"))
lat1[1:5]
```

```
## [1] 40.08462 40.12710 40.10810 40.12060 40.09360
```

```
lon1[1:5]
```

```
## [1] -88.19943 -88.22700 -88.21370 -88.20760 -88.20700
```

```
r = 6371000
lat2 = rep(40.1106, length(lat1))
lat_dif = (lat2-lat1)* (pi/180)
lon2 =rep(-88.2073, length(lon1))
lon_dif = (lon2-lon1) * (pi/180)
a =(sin(lat_dif/2)*sin(lat_dif/2))+cos(lat2)*cos(lat1)*(sin(lon_dif/2)*sin(lon_dif/2))
c= 2*atan2(sqrt(a),sqrt(1-a))
d=c*r/1609
urb_chm$distance2downtown=d
```

```
a = c(summary(urb_chm$distance2downtown),sd(urb_chm$distance2downtown, na.rm = TRUE))
names(a) = c(names(summary(urb_chm$distance2downtown)),"std")
a
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.      NA's       std
## 0.0876561 0.4993551 0.7204054 0.8834661 1.2194539 2.8044869 1.0000000 0.5118661
```

```
library(readr)
cham_county <- read_delim("https://uofi.box.com/shared/static/lbah1aw0cwghyrelzur0tk3akh65tin6.txt", "|
```

```
cham_county_2 <- cham_county %>%
  mutate(booking_datetime = as.POSIXct(paste(booking_date, booking_time))) %>%
  mutate(released_datetime = as.POSIXct(paste(replace_na(released_date, as.Date("2020-08-01")), release
  mutate(years_jail = as.numeric(difftime(released_datetime, booking_datetime, units = "days")) %/% 365
  mutate(days_jail = floor(as.numeric(difftime(released_datetime, booking_datetime, units = "days")) %%
  mutate(hours_jail = floor(as.numeric(difftime(released_datetime, booking_datetime, units = "hours")) %
  mutate(minutes_jail = floor(as.numeric(difftime(released_datetime, booking_datetime, units = "mins"))
  mutate(seconds_jail = floor(as.numeric(difftime(released_datetime, booking_datetime, units = "secs"))
  mutate(time_spent_in_jail1 = years_jail*365*24 + days_jail*24 + hours_jail + minutes_jail/60 + second
  mutate(time_spent_in_jail2 = as.numeric(difftime(released_datetime, booking_datetime, units = "hours"
```

```
result = cham_county_2 %>%
  select(c(`jacket_number`, `booking_date`, `booking_time`, `released_date`, `released_time`, `time_spen
```

```
cc <- cham_county %>%
  filter(city == "CHAMPAIGN" | city == "RANTOUL" | city == "SAVOY" | city == "URBANA") %>%
  select(jacket_number, `age arrest`, city, state, sex, race)
cc$city = sort(cc$city, decreasing = TRUE)
```

```
bp = cham_county_2[cham_county_2$race == 'Black',]
nbp = cham_county_2[cham_county_2$race != 'Black',]
mean(bp$time_spent_in_jail1, na.rm=TRUE)
```

```
## [1] 523.8519
```

```
mean(nbp$time_spent_in_jail1, na.rm = TRUE)
```

```
## [1] 358.2151
```