### Project Phase 1

### Due 15 June 2020

options(tidyverse.quiet = TRUE)  
library(plyr)  
library(tidyverse)  
library(lubridate)

library(GGally)

library(ggplot2)  
library(ggrepel)  
chicago <- read\_csv("chicago2.csv")

## Warning: Missing column names filled in: 'X1' [1]

## Parsed with column specification:  
## cols(  
## .default = col\_character(),  
## X1 = col\_double(),  
## ID = col\_double(),  
## Arrest = col\_logical(),  
## Domestic = col\_logical(),  
## Ward = col\_double(),  
## `Community Area` = col\_double(),  
## `X Coordinate` = col\_double(),  
## `Y Coordinate` = col\_double(),  
## Year = col\_double(),  
## Latitude = col\_double(),  
## Longitude = col\_double()  
## )

## See spec(...) for full column specifications.

# View(chicago)

#### Delete Columns, Mutate to Factors, Rename Columns

chicago <- read\_csv("chicago2.csv")

## Warning: Missing column names filled in: 'X1' [1]

## Parsed with column specification:  
## cols(  
## .default = col\_character(),  
## X1 = col\_double(),  
## ID = col\_double(),  
## Arrest = col\_logical(),  
## Domestic = col\_logical(),  
## Ward = col\_double(),  
## `Community Area` = col\_double(),  
## `X Coordinate` = col\_double(),  
## `Y Coordinate` = col\_double(),  
## Year = col\_double(),  
## Latitude = col\_double(),  
## Longitude = col\_double()  
## )

## See spec(...) for full column specifications.

# delete columns as instructed  
chicago <- chicago %>%  
 select(-X1, -ID, -`Case Number`, -`Updated On`, -`X Coordinate`, -`Y Coordinate`, -Location) %>%  
# convert to factors  
 mutate(Arrest = as.factor(Arrest)) %>%  
 mutate(Arrest = fct\_recode(Arrest, "No" = "FALSE", "Yes" = "TRUE")) %>%  
 mutate(Domestic = as.factor(Domestic)) %>%  
 mutate(Domestic = fct\_recode(Domestic, "No" = "FALSE", "Yes" = "TRUE")) %>%  
 mutate(`Primary Type` = as.factor(`Primary Type`)) %>%  
 mutate(`FBI Code` = as.factor(`FBI Code`)) %>%  
 mutate(Ward = as.factor(Ward)) %>%  
 mutate(`Community Area` = as.factor(`Community Area`)) %>%  
# convert the date obeject to R date/time object  
 mutate(Date = mdy\_hms(Date)) %>%  
# create new vriable of hour  
 mutate(hour = hour(Date))   
# rename columns for ease of typing  
 names(chicago)[names(chicago) == "Primary Type"] <- "PrimaryType"  
 names(chicago)[names(chicago) == "FBI Code"] <- "FBICode"  
summary(chicago)

## Date Block IUCR   
## Min. :2018-01-01 00:00:00 Length:15000 Length:15000   
## 1st Qu.:2018-04-11 22:23:00 Class :character Class :character   
## Median :2018-07-05 14:30:00 Mode :character Mode :character   
## Mean :2018-07-04 05:10:47   
## 3rd Qu.:2018-09-26 09:48:30   
## Max. :2018-12-30 23:40:00   
##   
## PrimaryType Description Location Description Arrest   
## THEFT :3616 Length:15000 Length:15000 No :11980   
## BATTERY :2784 Class :character Class :character Yes: 3020   
## CRIMINAL DAMAGE :1560 Mode :character Mode :character   
## ASSAULT :1131   
## DECEPTIVE PRACTICE:1077   
## OTHER OFFENSE : 995   
## (Other) :3837   
## Domestic Beat District Ward   
## No :12550 Length:15000 Length:15000 42 : 993   
## Yes: 2450 Class :character Class :character 28 : 714   
## Mode :character Mode :character 24 : 664   
## 27 : 634   
## 2 : 563   
## (Other):11431   
## NA's : 1   
## Community Area FBICode Year Latitude Longitude   
## 25 : 884 06 :3616 Min. :2018 Min. :41.64 Min. :-87.93   
## 8 : 709 08B :2333 1st Qu.:2018 1st Qu.:41.77 1st Qu.:-87.71   
## 32 : 634 14 :1560 Median :2018 Median :41.87 Median :-87.66   
## 29 : 516 26 :1402 Mean :2018 Mean :41.84 Mean :-87.67   
## 28 : 508 11 : 980 3rd Qu.:2018 3rd Qu.:41.91 3rd Qu.:-87.63   
## 43 : 461 18 : 803 Max. :2018 Max. :42.02 Max. :-87.53   
## (Other):11288 (Other):4306 NA's :252 NA's :252   
## hour   
## Min. : 0.00   
## 1st Qu.: 9.00   
## Median :14.00   
## Mean :13.04   
## 3rd Qu.:18.00   
## Max. :23.00   
##

### Look at the Data

#### Look at Primary Type, Ward, and Beat variables with largest numbers of observations

countPrimaryType <- (chicago) %>%  
 group\_by(chicago$PrimaryType) %>%  
 dplyr::summarise(Count=n()) %>%  
 mutate(Percent = round((Count/sum(Count)\*100))) %>%  
 arrange(desc(Count))%>%  
 filter(Count > 516 )  
# View (countPrimaryType)  
countPrimaryType

## # A tibble: 9 x 3  
## `chicago$PrimaryType` Count Percent  
## <fct> <int> <dbl>  
## 1 THEFT 3616 24  
## 2 BATTERY 2784 19  
## 3 CRIMINAL DAMAGE 1560 10  
## 4 ASSAULT 1131 8  
## 5 DECEPTIVE PRACTICE 1077 7  
## 6 OTHER OFFENSE 995 7  
## 7 NARCOTICS 803 5  
## 8 BURGLARY 657 4  
## 9 MOTOR VEHICLE THEFT 571 4

countWard <- (chicago) %>%  
 group\_by(chicago$Ward)%>%  
 dplyr::summarize(Count=n()) %>%  
 filter(Count > 400)

## Warning: Factor `chicago$Ward` contains implicit NA, consider using  
## `forcats::fct\_explicit\_na`

# View (countWard)  
countWard

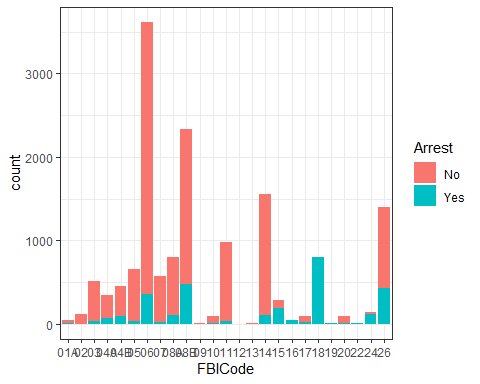
## # A tibble: 11 x 2  
## `chicago$Ward` Count  
## <fct> <int>  
## 1 2 563  
## 2 3 416  
## 3 6 509  
## 4 9 424  
## 5 17 473  
## 6 20 452  
## 7 21 465  
## 8 24 664  
## 9 27 634  
## 10 28 714  
## 11 42 993

countBeat <- (chicago) %>%  
 group\_by(chicago$Beat)%>%  
 dplyr::summarize(Count=n()) %>%  
 mutate(Percent = round((Count/sum(Count)\*100))) %>%  
 arrange(desc(Count))%>%  
 filter(Count > 110)  
# View (countBeat)  
countBeat

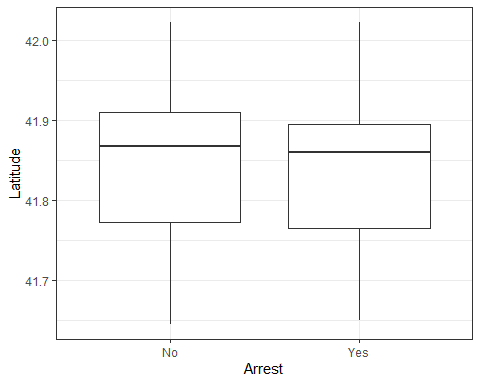
## # A tibble: 9 x 3  
## `chicago$Beat` Count Percent  
## <chr> <int> <dbl>  
## 1 1834 201 1  
## 2 0112 153 1  
## 3 0111 146 1  
## 4 0511 140 1  
## 5 0421 125 1  
## 6 1112 119 1  
## 7 1831 118 1  
## 8 1533 113 1  
## 9 0423 112 1

#### Look at Relationships between Variables

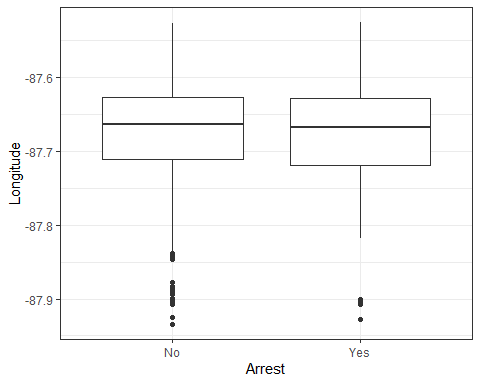
ggplot(chicago, aes(x=FBICode, fill=Arrest)) +  
 geom\_bar()+  
 theme\_bw()



# ggplot(chicago, aes(x=FBICode, fill=Arrest)) +  
# geom\_bar(position=position\_dodge())+  
# theme\_bw()  
  
ggplot(chicago, aes(x=Arrest, y=Latitude )) + geom\_boxplot() + theme\_bw()



ggplot(chicago, aes(x=Arrest, y=Longitude )) + geom\_boxplot() + theme\_bw()



#### Set Up for HeatMaps

#library(ggplot2)  
#library(plyr)  
#library(dplyr)  
chicago <- read\_csv("chicago2.csv")

## Warning: Missing column names filled in: 'X1' [1]

## Parsed with column specification:  
## cols(  
## .default = col\_character(),  
## X1 = col\_double(),  
## ID = col\_double(),  
## Arrest = col\_logical(),  
## Domestic = col\_logical(),  
## Ward = col\_double(),  
## `Community Area` = col\_double(),  
## `X Coordinate` = col\_double(),  
## `Y Coordinate` = col\_double(),  
## Year = col\_double(),  
## Latitude = col\_double(),  
## Longitude = col\_double()  
## )

## See spec(...) for full column specifications.

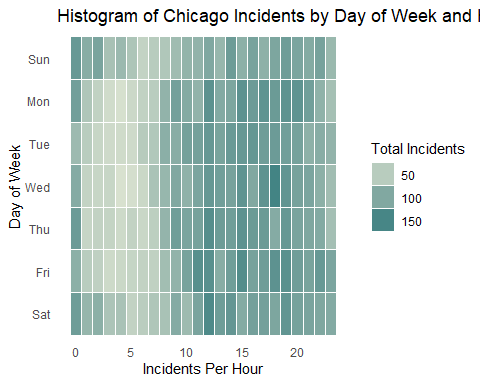
# chicago2 <- read\_csv("chicago2.csv", col\_types = cols(Date = col\_datetime(format = "%m/%d/%Y %I:%M:%S %p")))   
# Create variables for month year day-of-week and hour   
chicago$ymd <- mdy\_hms(chicago$Date)  
chicago$month <- month(chicago$ymd, label = TRUE)  
chicago$year <- year(chicago$ymd)  
chicago$wday <- wday(chicago$ymd, label = TRUE)  
chicago$hour <- hour(chicago$ymd)  
attach(chicago)  
  
# Assign color variables  
col1 = "#d8e1cf"   
col2 = "#438484"

#### Heat Map for Day and Hour

# Heatmap for Day and Hour  
dayHour <- ddply(chicago, c( "hour", "wday"), summarise,  
 N = length(ymd) )  
#reverse order of months for easier graphing  
dayHour$wday <- factor(dayHour$wday, levels=rev(levels(dayHour$wday)))  
attach(dayHour)

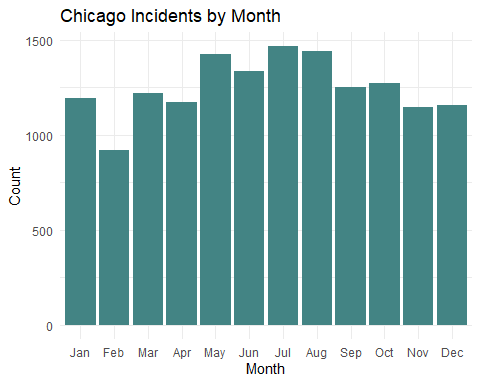
## The following objects are masked from chicago:  
##   
## hour, wday

ggplot(dayHour, aes(hour, wday)) + geom\_tile(aes(fill = N),colour = "white", na.rm = TRUE) +  
 scale\_fill\_gradient(low = col1, high = col2) +   
 guides(fill=guide\_legend(title="Total Incidents")) +  
 theme\_bw() + theme\_minimal() +   
 labs(title = "Histogram of Chicago Incidents by Day of Week and Hour",  
 x = "Incidents Per Hour", y = "Day of Week") +  
 theme(panel.grid.major = element\_blank(), panel.grid.minor = element\_blank())



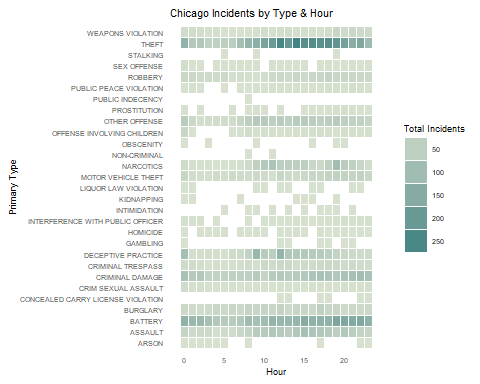
#### Bar Chart for Month

# Barchart for Month  
ggplot(chicago, aes(x=month)) + geom\_bar(fill=col2) +  
 scale\_fill\_gradient(low = col1, high = col2) +   
   
 labs(title = "Chicago Incidents by Month ",  
 x = "Month", y = "Count") +  
 theme\_bw() + theme\_minimal()



#### Incidents by Type and Hour

names(chicago)[names(chicago) == "Primary Type"] <- "PrimaryType"  
groupSummary <- ddply(chicago, c( "PrimaryType", "hour"), summarise,  
 N = length(ymd) )  
#overall summary  
ggplot(groupSummary, aes( hour,PrimaryType)) + geom\_tile(aes(fill = N),colour = "white") +  
 scale\_fill\_gradient(low = col1, high = col2) +   
 guides(fill=guide\_legend(title="Total Incidents")) +  
 labs(title = "Chicago Incidents by Type & Hour",  
 x = "Hour", y = "Primary Type") +  
 theme\_bw() + theme\_minimal() +   
 theme(panel.grid.major = element\_blank(), panel.grid.minor = element\_blank()) +  
 theme(text = element\_text(size=7))



#### Incident Locations

library("ggmap")

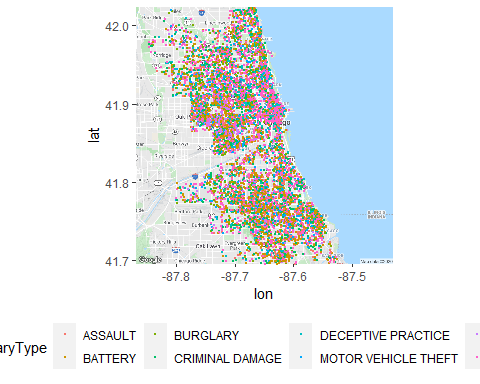
## Google's Terms of Service: https://cloud.google.com/maps-platform/terms/.

## Please cite ggmap if you use it! See citation("ggmap") for details.

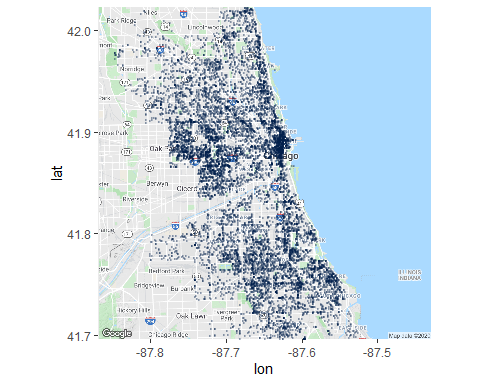
ggmap::register\_google(key = "AIzaSyAcyuuUd02G1743CLVML71ROQZaqCmjZcQ")  
  
col1 = "#011f4b"  
  
chicagoMap <- chicago %>%  
 filter( PrimaryType == "THEFT"| PrimaryType =="BATTERY" |   
 PrimaryType =="CRIMINAL DAMAGE" | PrimaryType =="ASSAULT" |   
 PrimaryType =="DECEPTIVE PRACTICE" | PrimaryType =="NARCOTICS" |  
 PrimaryType =="BURGLARY" | PrimaryType =="MOTOR VEHICLE THEFT" )  
p <- ggmap(get\_googlemap(center = c(lon = -87.65, lat = 41.86),  
 zoom = 11, scale = 2,  
 maptype ='terrain',  
 color = 'color'))

## Source : https://maps.googleapis.com/maps/api/staticmap?center=41.86,-87.65&zoom=11&size=640x640&scale=2&maptype=terrain&key=xxx

p + geom\_point(aes(x = Longitude, y = Latitude, colour = PrimaryType), data = chicagoMap, size = 0.5) +   
 theme(legend.position="bottom")



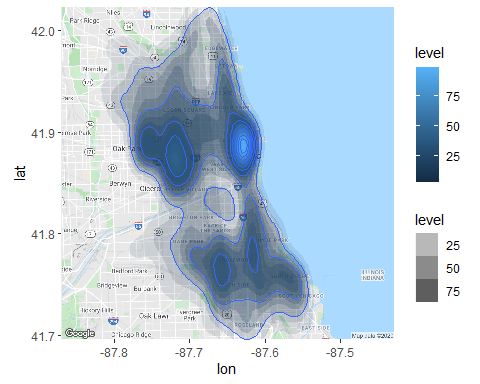
p + geom\_point(aes(x = Longitude, y = Latitude), colour = col1, data = chicagoMap, alpha=0.25, size = 0.5) +   
 theme(legend.position="none")



library(ggrepel)

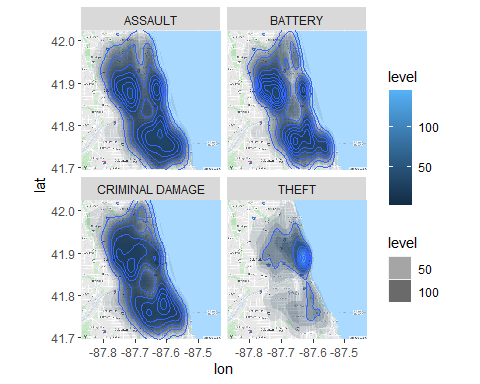
#### Density Map, All Primary Types

p + stat\_density2d(  
 aes(x = Longitude, y = Latitude, fill = ..level.., alpha =..level..),  
 size = 0.2, bins = 30, data = chicagoMap,  
 geom = "polygon"  
 ) +  
 geom\_density2d(data = chicagoMap,   
 aes(x = Longitude, y = Latitude), size = 0.3)



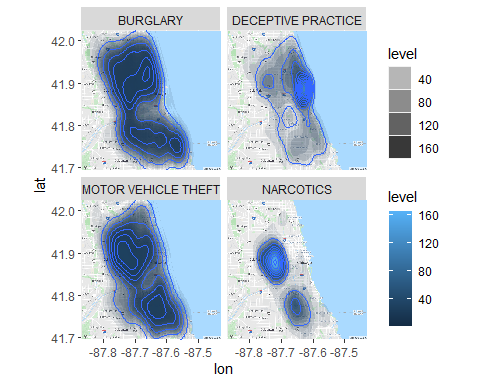
#### Density Map, Four Most Frequent Primary Types

chicagoMap2 <- chicago %>%  
 filter( PrimaryType == "THEFT"| PrimaryType =="BATTERY" |   
 PrimaryType =="CRIMINAL DAMAGE" | PrimaryType =="ASSAULT" )   
  
p + stat\_density2d(  
 aes(x = Longitude, y = Latitude, fill = ..level.., alpha =..level..),  
 size = 0.2, bins = 30, data = chicagoMap2,  
 geom = "polygon"  
 ) +  
 geom\_density2d(data = chicagoMap2,   
 aes(x = Longitude, y = Latitude), size = 0.3) +  
 facet\_wrap(~ PrimaryType, nrow=2)



#### Density Map, 4th thru 8th Most Frequent Primary Types

chicagoMap3 <- chicago %>%  
 filter( PrimaryType =="DECEPTIVE PRACTICE" | PrimaryType =="NARCOTICS" |  
 PrimaryType =="BURGLARY" | PrimaryType =="MOTOR VEHICLE THEFT" )   
  
p + stat\_density2d(  
 aes(x = Longitude, y = Latitude, fill = ..level.., alpha =..level..),  
 size = 0.2, bins = 30, data = chicagoMap3,  
 geom = "polygon"  
 ) +  
 geom\_density2d(data = chicagoMap3,   
 aes(x = Longitude, y = Latitude), size = 0.3) +  
 facet\_wrap(~ PrimaryType, nrow=2)



#### Density Map, All Types, Black Points

p + geom\_point(aes(x = Longitude, y = Latitude), colour = "black", data = chicagoMap, alpha=0.3, size = 0.5) +   
 theme(legend.position="none")

