# **Final Report**

Introduction

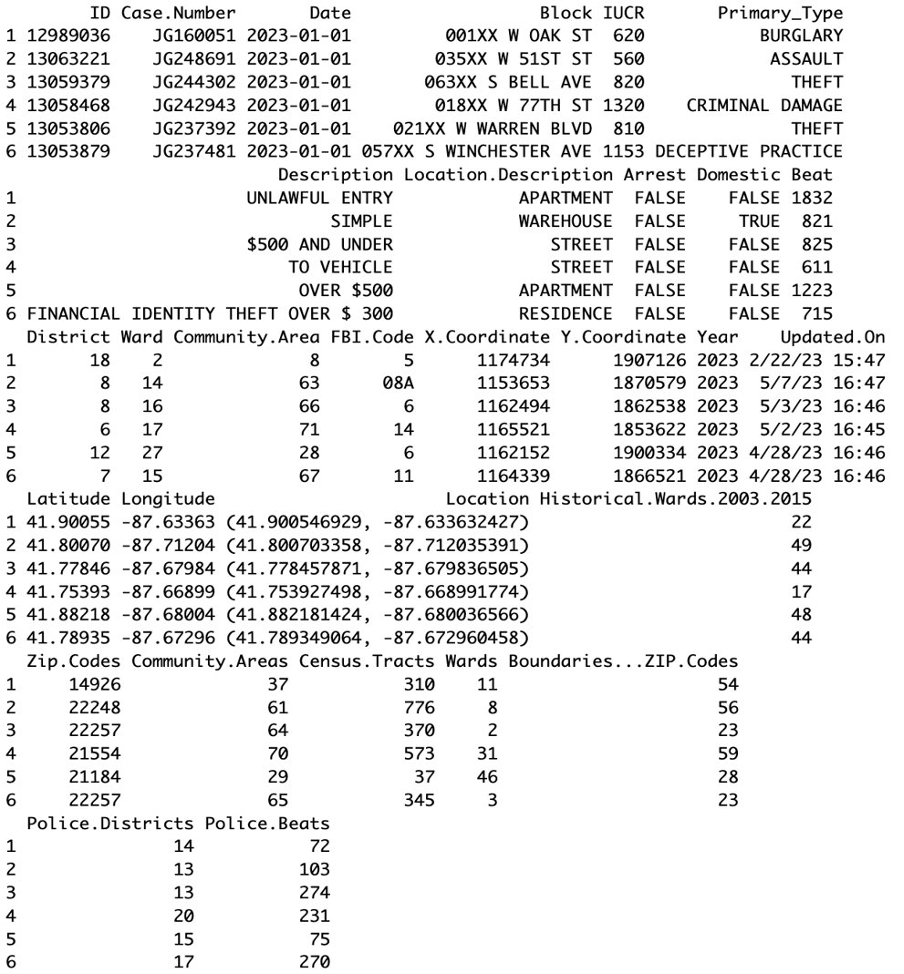
Chicago Crime Dataset

Source: <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2>

We are working on Chicago Crimes dataset. Our dataset contains 1,048,575 observations with 30 variables which provides comprehensive information on criminal incidents reported in Chicago, Illinois. It offers valuable insights in analyzing crime patterns and also helps in evidence-based decision making. Our data encompasses a wide range of attributes including incident details, location, time, classification and many more.

Initially, this was a huge dataset with about 7 million observations i.e., starting from the year 2001 until present. As it was a huge data, we took a sample of it for the past few years i.e., 2018 to 2023. Our variables include ID, Case number which shows details regarding the cases that have been recorded, then there is the type of the crime along with its description and location where the crime occurred. In addition to this there are details regarding beats, districts, zipcodes etc.

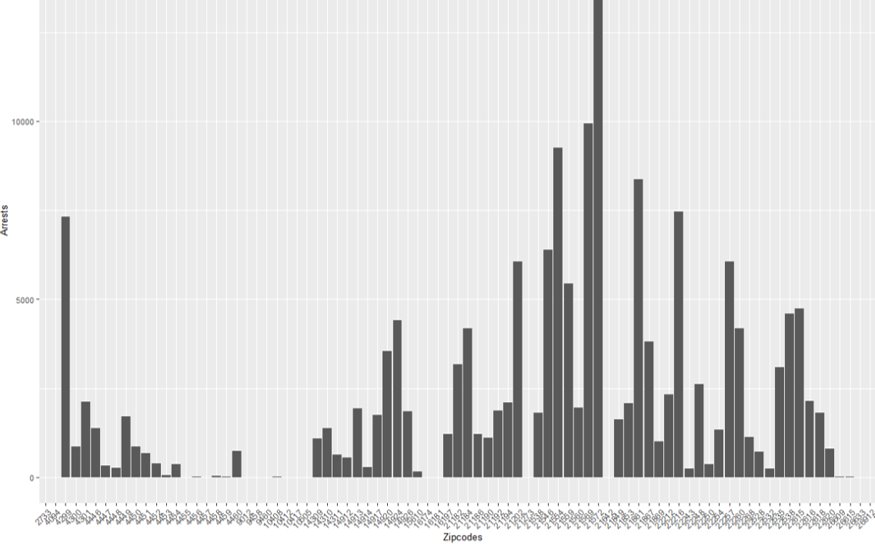
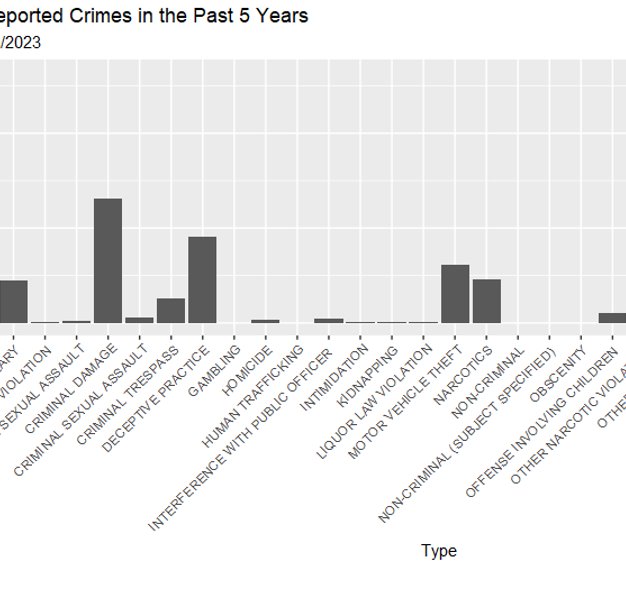
We are trying to visualize what locations had most common types of crimes and along with it we are also trying to see if the arrests have been made or not for each type of crime that have been recorded. Below are the variables that we will be working with for our visualizations:



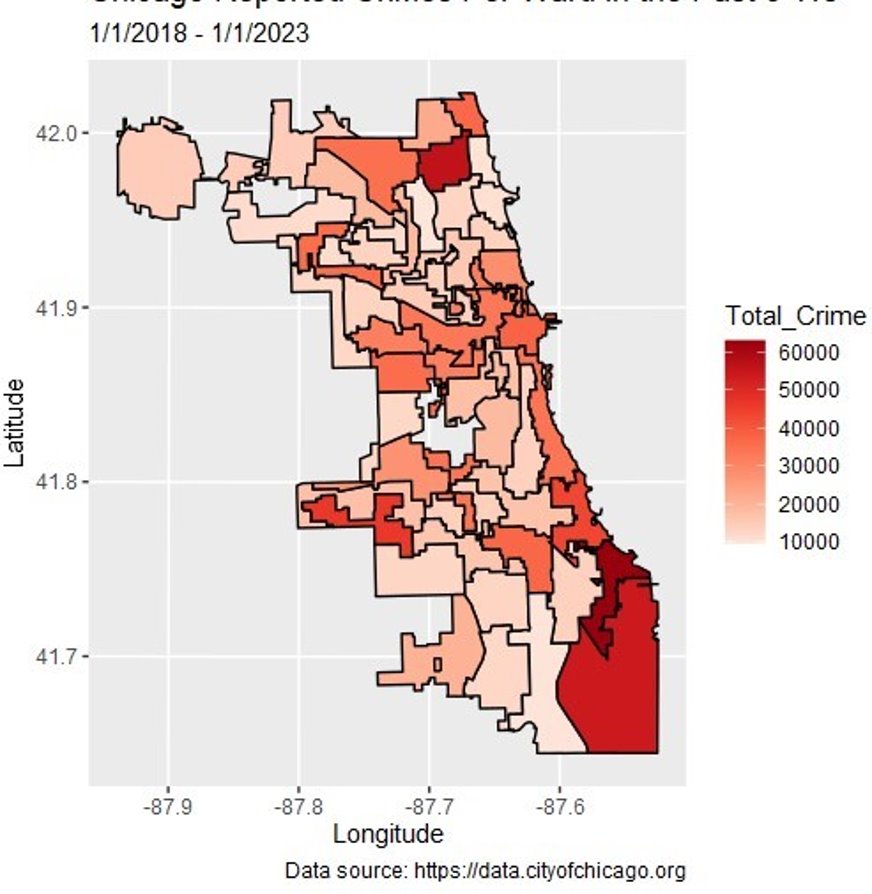
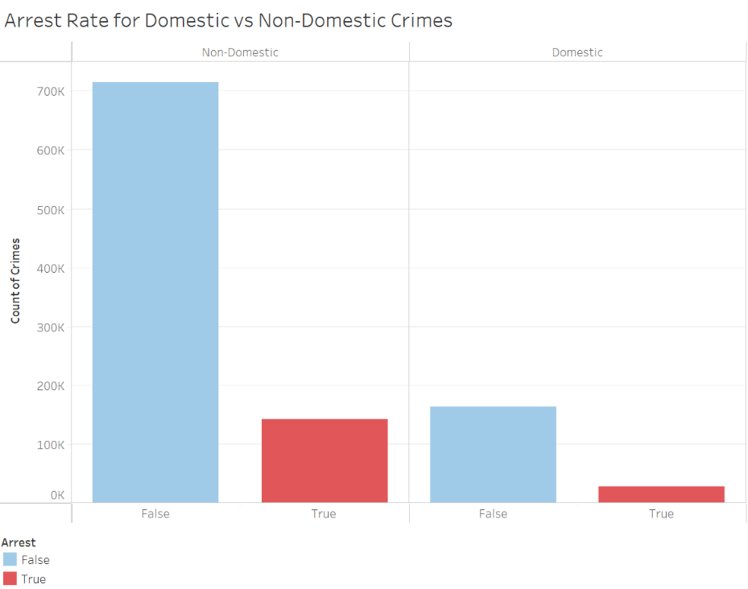
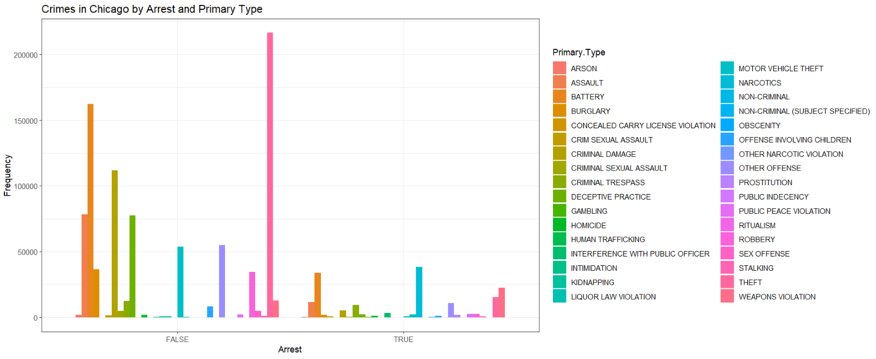
# **Exploratory Analysis**

To start with the exploratory analysis, we began by looking at the most common variables that would make a difference when it came to crime data. We know Chicago has a high crime rate so there were several pieces we would want to focus on when it comes to the initial understanding of the data and what visualizations would make the most sense. The variables we ended up choosing to focus on for creating visualizations were Crime Type, Arrest, Location, and Building Type/Street. The type of crime is important as it helps us understand the distribution of such a large dataset of Chicago crime history. Not all crimes lead to an arrest, this makes it important to understand the ratios for different crimes and how often it was reported versus an arrest being made. Location plays a big part as well; we all know certain neighborhoods may be more dangerous than others so those should stick out. With a variety of location data such as zip code, wards, police districts/beats, and latitude/longitude there are a lot of options to explore. The general building and street locations can also help to show hotspots for different crimes like out in the street or in an apartment or at certain intersections that may want some more attention.

Some of the visualizations we ended up creating focused on counts of crime, and crime based on locations. There were not complete visualizations but gave us an idea of what we wanted to continue to explore and create more complex iterations of visualizations for later milestones and final report.



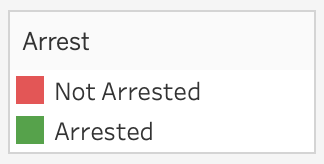
Another focus was arresting rates for certain crimes. Some crimes can be less severe or just be harder to catch so come with much lower arrest rates than others that are reported from the data.

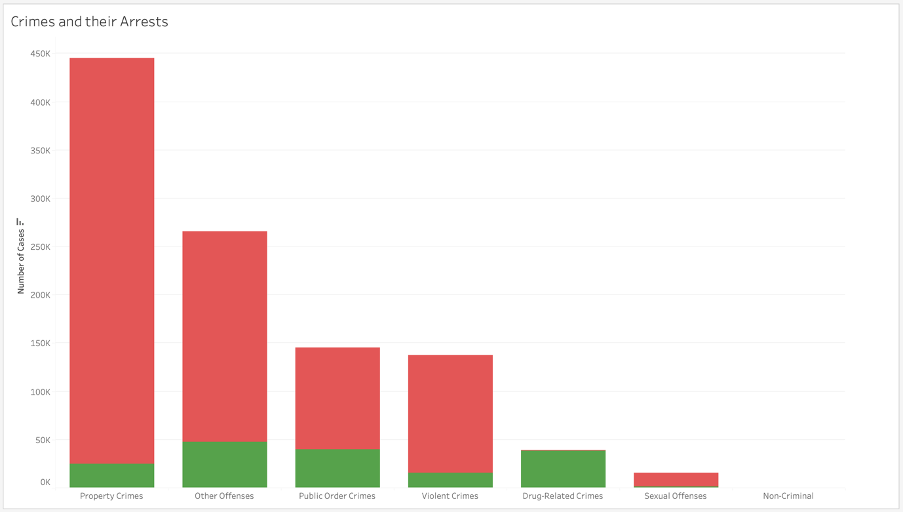
Locations were going to be a very important factor when it comes to crimes and where they are more common. One of the more advanced visualizations we started with was a choropleth as it was an easy way to see the entire Chicago area. By using a map file, we were able to attribute crimes to certain locations and pick out the more dangerous areas of the city. This could be done at various levels based on the location data that was provided with the crimes like ward, zip-code, and longitudinal.

These were our starting out visualizations that helped us pick on where we wanted to focus our efforts with all the data, we had available and the trends we wanted to try and highlight. For this reason, you will see many of the completed visualizations are built off the initial exploratory versions that were first started.

# Visualizations

STACKED BAR CHART:





The visualization here is a stacked bar chart which shows the types of crime that have been recorded and if the arrests have been made or not with the y-axis representing the count of cases and the x-axis displaying the types of crimes. The bars are stacked to show the total count of cases for each crime type. The graph here provides an overview of the distribution of crime types and allows for a comparison of arrest rates across different types of crime.

I started my visualization by grouping the related types of crimes into various groups.

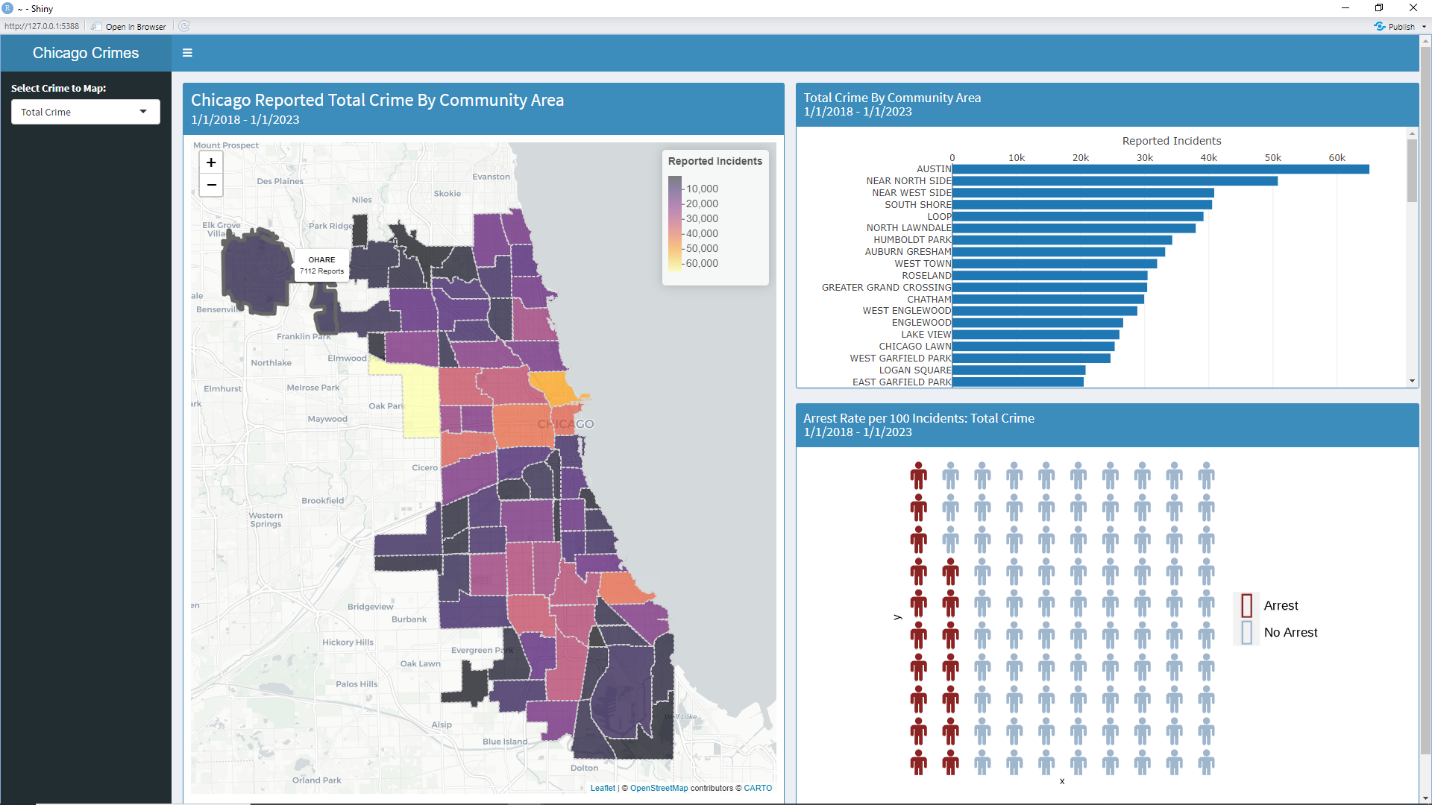
1. Property Crimes: Arson, Burglary, Criminal damage, Motor vehicle theft, Theft
2. Violent Crimes: Assault, Homicide, Kidnapping, Robbery, Sex offense, Stalking
3. Public Order Crimes: Concealed carry license violation, Criminal trespass, Deceptive practice, Gambling, Liquor law violation, Prostitution, Public indecency, Public peace violation, weapons violation
4. Sexual Offenses: Criminal sexual assault, offense involving children
5. Drug Related Crimes: Narcotics, Other narcotic violation
6. Non Criminal
7. Other Offenses: Battery, Human trafficking, Interference with public officer, Intimidation, Obscenity, Ritualism

Plotting these groups on x-axis and considering the number of cases reported on y-axis. I have also included color scheme to for arrest rate as to see if the arrest have been made or not. Red color indicates “Not Arrested” and green indicates “Arrested”. In addition to this, I have ordered the x-axis in descending order for better understanding the relationship between the crimes and the number of cases.

From this visualization it can be said that Property Crimes are one of those which had highest number of cases recorded, then comes Other Offenses followed by Public Order Crimes and so on. One noticeable thing in here is that most of the suspects for these crimes have not been arrested, very few of them are arrested. But in case of Drug Related Crimes arrest rate is positive as compared to those not arrested for that particular crime out of the total number of records for drug crimes.

Overall, we can say that Property Crimes are the highest that’s being recorded over the years in Chicago and very less suspects have been arrested. Among all these crimes Drug Related Crimes are the only kind which had better arrest rates i.e., most of the suspects have been arrested compared to the others.

Interactive Dashboard:



**Visualization Description:**

This interactive dashboard consists of three components: a choropleth, a bar chart, and an isotype plot. The interactive component comes from the user of the application being able to select which type of crime they wish to analyze, as well as both the choropleth and bar chart displaying text widgets when hovered over. The choropleth and bar chart both depict the frequency of reported incidents of a given crime by community area, while the isotype plot shows the arrest rate.

**Drafting Process:**

The original plan for this visualization was just a choropleth, without any interactivity. Since choropleths have an issue of larger shapes potentially being overemphasized, a diffusion cartogram was considered - mapping the crime incidents to size as well. This helped with the overemphasis on size, however it seemed to muddy the message of the geospatial distribution of crimes in Chicago. While Chicago is a well-known city, its general layout is not nearly as well-known as something like the United States. This resulted in the distortion created by the cartogram making the map difficult to contextualize. Finally, based on some feedback, a background map was added to the choropleth. This was done by using leaflet instead of ggplot2, which was what was done originally. This also added the functionality of zooming and panning the map, providing an even wider audience with a better perspective on location.

Once the choropleth was satisfactory, a bar chart was created to help the audience understand the values that were also mapped on the choropleth. The choropleth is useful for gaining insights into a general trend; however, color is not a great encoding for distinguishing specific values. The sorted bar chart makes it so that the number of reported incidents is much easier to read and compare between community areas. Finally, the bar chart was recreated using plotly instead of ggplot2, which added a text popup functionality. This made the values even more readily available for the audience.

Since there were now multiple visualizations, a dashboard seemed appropriate. Originally, creating a grid of choropleths of different crimes had been considered. Since Shinydashboard allows for interactivity, this idea was brought back in the form of creating a user input in which the crime to map could be selected. This allowed the audience to compare the distributions of different crimes. Due to the visualization now cycling through crimes, as well as our group’s desire to analyze arrests, a visualization was added that would map the arrest rates of the user selected crime.

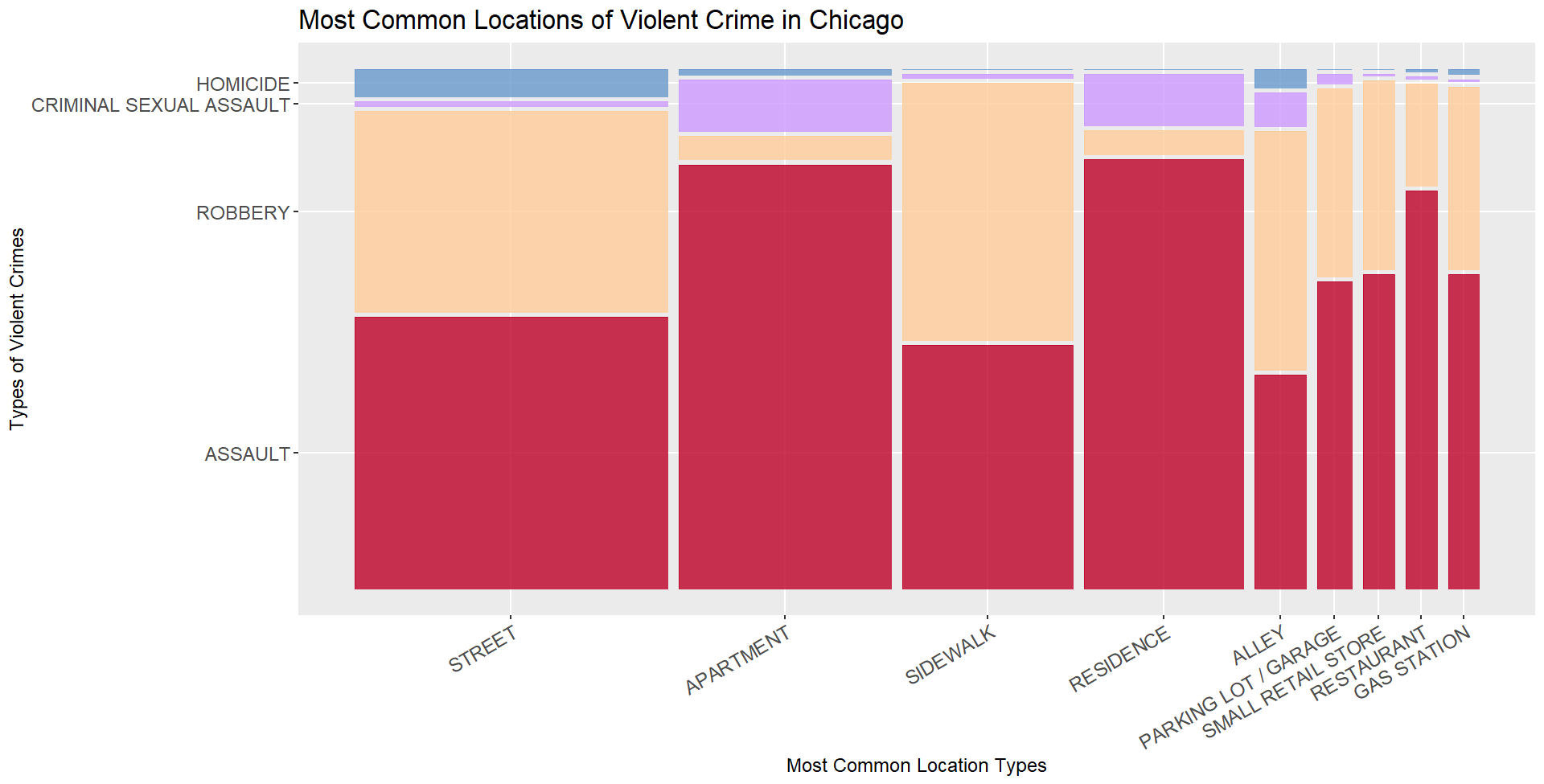
To visualize arrest rates, initially a pie or donut chart was considered, however those are generally considered bad practice, as humans aren’t particularly good at estimating quantity from angles. The chart would represent two parts of a whole: incidents resulting in an ‘arrest’ or ‘no arrest’. This seemed appropriate for a waffle plot, since humans are better at decoding areas. Once this was created, basic glyphs were added to help with short term memorability. This made the visualization into an isotype plot.

**Story:**

The story that this visualization shows is that crimes are generally much more prevalent in the downtown, west, southwest, and far south sides of Chicago. Austin, Near North Side, as well as Near West Side are particularly active for most crime types. In terms of arrest rates, the analysis shows that they vary wildly from crime to crime. One potential interpretation is that it greatly depends on the way certain crimes are known to have occurred. How a theft may be reported versus how a weapons violation may be reported is likely quite different. This might account for the vastly different arrest rates.

Mosaic Plot

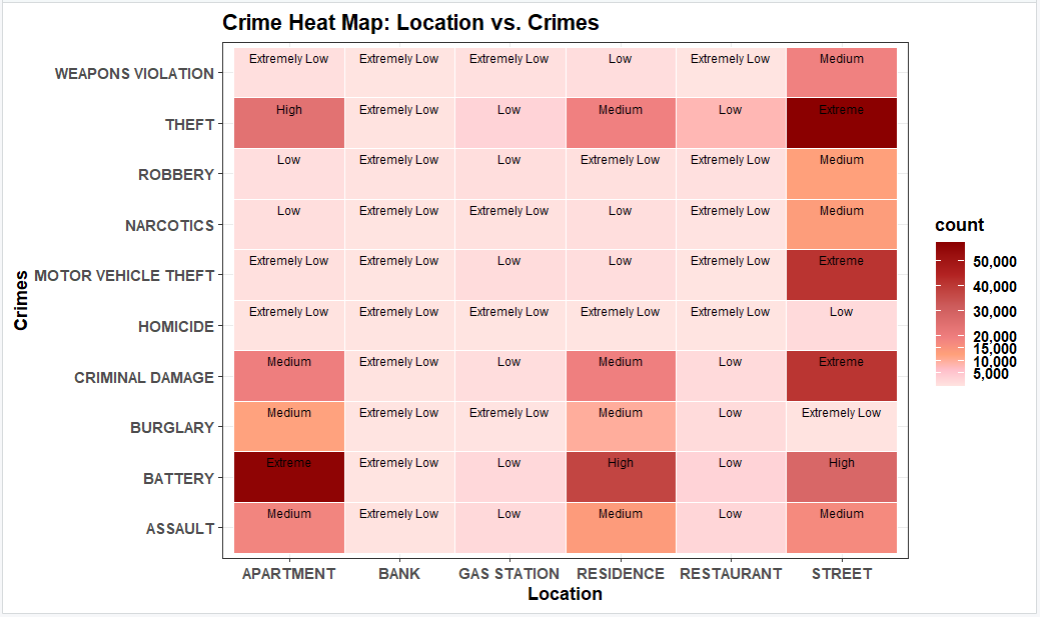
The following visualization is a Mosaic plot depicting the four types of violent crimes and the top nine locations where they are most frequently reported. The choice to use the four types of crimes was based on the FBI's classification of violent crimes. Mosaic plots offer a quick overview of the distribution of the data, using area and color to make clear distinctions. The benefit of such a plot is that it makes the most prevalent data points immediately clear. In this case, it is evident that assault and robbery make up the largest sections of the reported violent crimes. In addition, we can see that the highest density of reported crimes takes place in street settings. While the mosaic plot does not offer a look at the fine details or exact values, it excels at providing viewers with a high-level summary.



# Analysis

Knowing that Chicago is a high crime city we were interested in exploring the most reported issues both in terms of crime type and locations. Our visualizations assessed a few different variables within the data and allowed us to identify a couple key trends. In our exploratory visualizations we could see that certain areas had much higher crime rates than others so we developed a visual in which this could be more easily seen. Using this to look at the specific regions of Chicago we found that the areas with the most reported crimes are Austin, the area Near North Side, as well as Near West Side. It was also evident that particular location types, such as streets or stores were also more frequent so we worked on visuals that would provide greater detail on this aspect. Our analysis showed that street settings are the most common location of report crimes, overall. In addition, we found that certain crimes were more frequent in specific locations. For instance, battery is most common in apartments and theft is most common in street settings. Focusing exclusively on violent crimes our visuals demonstrate that assault and robbery make up the majority. Finally, incorporating the time aspect demonstrated that crime counts have stayed somewhat consistent, within 15% points of the previous year's counts. In contrast, we see that overall arrest counts have been exhibiting a slight decrease over the years. Taking a closer look at arrest rate, we have found that for every 100 reported crimes there were only 17 arrests.

**HEAT MAP VISUALIZATION:**



**Introduction**:

The crime heat map visualization provides a comprehensive overview of the relationship between different crime types and their corresponding locations. By visually representing crime data, this visualization aims to uncover patterns, trends, and insights that can assist in understanding crime distribution and inform decision-making processes.

**Data Description:**

The data used for this visualization consists of crime records that include information about the crime type and the location where it occurred. The dataset includes crimes such as theft, battery, criminal damage, assault, robbery, burglary, motor vehicle theft, narcotics, weapons violation, and homicide. The locations range from streets and residences to apartments, gas stations, restaurants, and banks.

**Exploratory Analysis:**

Using the data, I developed a heat map visualization to represent the crime data effectively. The heat map displayed different shades of color to indicate the crime counts, with darker colors representing higher crime rates and lighter colors representing lower crime rates. By analyzing the heat map, I was able to identify areas with higher crime rates and areas with lower crime rates, providing insights into crime distribution.

To enhance the understanding of the visualization, I included labels on the map to categorize the crime rates into different levels: "Extreme," "High," "Medium," "Low," and "Extremely Low." These labels helped me interpret the crime rates more meaningfully and identify areas with varying degrees of criminal activity.

**Conclusion**:

By creating the crime heat map visualization, I gained valuable insights into the spatial distribution of crimes and the patterns associated with different crime types. This information can be used to make informed decisions related to crime prevention, resource allocation, and policy development.

However, it's important to acknowledge that the visualization has some limitations. It relies on the available crime data and may not capture the complete picture of crime in the area. Factors such as unreported crimes or variations in data collection methods can influence the accuracy and representativeness of the visualization.

In summary, the crime heat map visualization provides a valuable tool for understanding crime patterns and identifying areas of concern. This information can be used to develop effective strategies for crime prevention and public safety improvement.

# Appendix 1: Individual Reports

### Jorge Fernandez:

**Summary of Work:**

In this project, my primary role was to create a visualization, as well as submit milestones. I also helped with overall planning and communication.

The visualization I created was the interactive dashboard. This was done on my own and took up much of my time. Learning how to use Shinydashboard as well as iteratively improving the charts in the dashboard was a long but worthwhile process. As described in the report, I initially planned to create just the choropleth, however feedback was provided that suggested pairing it with a bar chart. From there I added and edited the visualizations until I ended up with the submitted dashboard.

Since I oversaw submitting the milestones, I often communicated with the group about the planning and timing of submissions. Our group had much difficulty in finding times for all of us to meet, so most of the communication was done via text or discord. As we were unable to meet for the presentation, I was also in charge of receiving each team member’s individual recordings and editing them together.

**Summary of Takeaways:**

This course has taught me so much about the world of visualizations. Not only did I learn about types of charts I've never even heard of, but more importantly I learned about how to approach creating visualizations in general.

Before taking this course, I had a rather simplistic view of data visualization. The extent of my knowledge was just choosing the right chart type and plugging in the data. However, I now understand that the process is far more nuanced and complex. It starts with a deep understanding of the data at hand. This is why the initial step of creating exploratory visualizations is crucial. It helps to uncover patterns, trends, and outliers in the data that may not be immediately apparent.

I also learned about the importance of considering the audience and the context when creating a visualization. I learned that the best visualizations are those that are tailored to the specific needs and background of the audience. This requires an understanding of not just the data, but also the subject matter and the audience's prior knowledge and expectations.

I used this knowledge, as well as what was taught in R, in order to create the interactive dashboard. The long and iterative process of creating the dashboard reinforced what was taught in the class. Not only did I gain experience in creating visualizations in R using various packages, I also truly came to understand how important exploratory visualizations are in choosing a direction of analysis and executing it efficiently.

Alejandra Rodriguez:

My role within the team was to help organize and coordinate our efforts by keeping track of deadlines and scheduling meetings. For my visualization contributions I tried to focus on one of the less prominent aspects of our data. I originally started working with the domestic vs non-domestic variable. I plotted the arrest rates with respect to domestic and non-domestic crimes, but it did not produce any insightful revelations. I then shifted my focus to a particular type of crime, filtering our data specifically on crimes categorized as violent. This is what led to the mosaic plot of violent crime types and most common locations. I found that honing in on a particular segment of crimes provided an additional perspective in our analysis.

Throughout this course I learned a lot about different types of plots and their uses. I was surprised by the number of plots I had been previously unfamiliar with. Going hand in hand with all the new plots, I learned about the various uses for different types of plots and came to understand why certain visualizations are much better suited to particular data than others. I was also pleasantly surprised by how much I learned about human perception and encodings. The psychology behind perception and explanation of encoding types was very valuable when it came to choosing plot types. In addition, learning about the ways in which a plot can be distorted continues to be very helpful in assessing the reliability of a visual. Most importantly, I gained a lot of knowledge in R. My previous R experience did not involve anything beyond a basic plot so the tutorials provided very useful information. Not only do I now feel a lot more comfortable in R, I have also learned about the best resources for figuring out how to do something new in R.

**MONISHA ARUNACHALAM**

**Story:**

During the initial stages of the project, I began by creating a bar chart to visualize the relationship between different crime types and the corresponding arrest rates. This allowed us to compare the arrest rates across various crimes and gain insights into the effectiveness of law enforcement in different areas. It provided a clear overview of the data and highlighted any discrepancies or patterns that emerged.

As we progressed to the next milestone, I decided to experiment with a **heat map visualization**. Heat maps offer a different perspective and can provide more insights when dealing with complex data. By mapping the crimes across different locations, we were able to uncover spatial patterns and identify areas with higher crime rates. The heat map allowed us to visualize the density of crimes in different areas and draw conclusions about hot spots or areas of concern.

Throughout the project, both my teammates and I worked on creating more sophisticated visualizations to enhance our understanding of the data. This process involved refining and iterating on our initial designs, incorporating feedback, and adjusting to ensure clarity and effectiveness. We learned the importance of selecting appropriate visualization techniques for different types of data and research questions.

I primarily **used R, for creating my visualizations.** R provided me with a flexible and powerful platform to analyze and visualize data. Working with R expanded my skills in data visualization and gave me hands-on experience in using programming languages for effective data analysis and visualization.

Through these experiences, I gained valuable insights into the power of data visualization and its ability to uncover hidden patterns and communicate information effectively. I developed a better understanding of the strengths and limitations of various visualization techniques, and the importance of iterative design and continuous improvement.

Overall, this project provided an excellent opportunity for me to explore and learn from a range of visualizations, from simpler bar charts to more complex heat maps. It deepened my understanding of data visualization principles and techniques, and I am now equipped with valuable skills that I can apply to future projects and analyses.

SANJANA GOWDA HETTHUR CHANDRASHEKAR:

I have been with this team in all sorts of decision making related to the data and the visualizations. I have also made my point wherever it’s necessary.

Initially I started with stacked bar chart for my visualization in order to visualize if the arrests have been made or not for different types of crimes. I chose stacked bar chart as it is one of those graphs that’s easily understandable and helps in comparing with other values. While doing this, I thought it would be interesting to check how these arrest rates have changed over the years and tried to visualize it using a line graph. It pretty much gave the expected result except that it was dropping for the current year as the data was only until Jan 2023. Then I continued working on stacked bar chart that was constructed initially by ordering and grouping it for better understanding. Throughout the process one surprising thing for me was for almost all the crimes it was about less 30% suspects have been arrested.

Data Visualization course has given me a new perspective to the data. Before taking this course I was familiar with few graphs but in this course I got to know that there’s a lot more and it’s possible to use them in the way we want to describe things. And it also taught me how even the smallest things in a graph matters a lot, for example the color encoding. In addition to this it also taught me what graphs should not be used for specific type of data. I got better with this while working on assignments, projects where I had the chance to apply things my way. I knew R, but this was the first time I was using Tableau. This course also introduced me to Tableau and how to work with it.

### **Saiyed Irfanullah**

My role within the team was to help organize and make sure we were on track with our deadlines, sending reminder communications as needed, also creating visualizations and helping team members with whatever else they may have needed. I created some of the initial visualizations when it came to location data and general crimes. I was particularly interested in zip codes and how crime would relate to different neighborhoods in the Chicago area. I started some choropleth visualizations when it came to showing crime on a map and got help from the team to complete it. I worked on coming up with ideas of what we should focus on for the story we wanted to present to the audience. I recorded and helped present the exploratory analysis sections for the presentation and report as those were areas where I had helped with multiple pieces and could help explain. I really enjoyed working with the large dataset and coming up with new ideas for how we could show visualizations that would help bring a story from the vast amount of crime, especially when it came to location related data.

This course has taught me a lot about visualizations on how important it is to get them correct so as to not show the wrong message to your audience. I really liked learning about certain graphs that can be misleading to normal people who do not have the knowledge to read them in depth. There were times I really struggled with creating visualizations where I thought I had everything figured out but for some reason my data has issues, I spent hours often troubleshooting these problems. As someone who does not use R very often this would be troublesome for me creating some of the choropleths definitely was a large learning experience just based on time for me to get it all right with data in the correct format. For my current jobs, I do create basic visualizations and interactive plots for business needs and this class has definitely given me more knowledge that can be implemented. A takeaway I want to build on in my professional career from here would be well-designed graphs like heatmaps and density plots to show trends that I could display.

# Appendix 2: Code & Exploration Results

Interactive Dashboard Code:

***Loading Libraries/Data & Cleaning/Processing:***

library(data.table)

library(ggplot2)

library(dplyr)

library(rgdal)

library(maptools)

library(sp)

library(sf)

library(viridis)

library(ggwaffle)

library(emojifont)

library(leaflet)

library(plotly)

# Function to clean some variable names

clean\_names <- function(.data, unique = FALSE) {

n <- if (is.data.frame(.data)) colnames(.data) else .data

n <- gsub("%+", "\_pct\_", n)

n <- gsub("\\$+", "\_dollars\_", n)

n <- gsub("\\++", "\_plus\_", n)

n <- gsub("-+", "\_minus\_", n)

n <- gsub("\\\*+", "\_star\_", n)

n <- gsub("#+", "\_cnt\_", n)

n <- gsub("&+", "\_and\_", n)

n <- gsub("@+", "\_at\_", n)

n <- gsub("[^a-zA-Z0-9\_]+", "\_", n)

n <- gsub("([A-Z][a-z])", "\_\\1", n)

n <- tolower(trimws(n))

n <- gsub("(^\_+|\_+$)", "", n)

n <- gsub("\_+", "\_", n)

if (unique) n <- make.unique(n, sep = "\_")

if (is.data.frame(.data)) {

colnames(.data) <- n

.data

} else {

n

}

}

# Read in data

ChiCrimes <- fread("C:/Users/Family/Downloads/Crimes\_-\_2018\_to\_2023.csv")

# Drop columns that I know I won't need at all

ChiCrimes <- ChiCrimes[, -c(25:27)]

#-------------------------------------------

# Create DF of Type and Location Columns

#-------------------------------------------

# get columns of interest

locationData <- ChiCrimes[, c('Primary Type', 'Arrest', 'Latitude', 'Longitude', 'Ward',

'Zip Codes', 'Community Area', 'District', 'Beat',

'Boundaries - ZIP Codes', 'Police Districts',

'Police Beats')]

# rename some columns

locationData <- locationData %>%

clean\_names()

locationData$zip\_boundaries <- locationData$boundaries\_minus\_zip\_codes

locationData <- locationData[,-c(10)]

# drop nas

locationData <- na.omit(locationData)

# convert to Dataframe

locationData <- as.data.frame(locationData)

#-------------------------------------------

# Get shapefile for choropleth

#-------------------------------------------

community\_areas <- readOGR(dsn = 'C:/Users/Family/OneDrive - DePaul University/DSC 465/Boundaries - Community Areas (current)',

layer = 'geo\_export\_91d3ab7f-61e5-4936-9005-86be2160d761')

#-------------------------------------------

# Group by and summarize crime types

#-------------------------------------------

crimes <- locationData %>%

group\_by(community\_area) %>%

summarize(`Total Crime` = n(),

Theft = sum(primary\_type == "THEFT"),

Battery = sum(primary\_type == "BATTERY"),

`Criminal Damage` = sum(primary\_type == "CRIMINAL DAMAGE"),

Assault = sum(primary\_type == "ASSAULT"),

`Deceptive Practice` = sum(primary\_type == "DECEPTIVE PRACTICE"),

`Vehicle Theft` = sum(primary\_type == "MOTOR VEHICLE THEFT"),

Narcotics = sum(primary\_type == "NARCOTICS"),

Burglary = sum(primary\_type == "BURGLARY"),

Robbery = sum(primary\_type == "ROBBERY"),

`Weapons Violation` = sum(primary\_type == "WEAPONS VIOLATION"))

# create column for merging

crimes$area\_num\_1 <- as.character(crimes$community\_area)

# Merge the spatial data with crimes data

merged\_data <- merge(community\_areas, crimes, by = "area\_num\_1")

# write to working directory for shinydashboard

write.csv(merged\_data, "output.csv", row.names=FALSE, quote=FALSE)

#-------------------------------------------

# Group by and summarize arrest data

#-------------------------------------------

arrestData <- locationData[ , c("primary\_type", "arrest")]

arrestData <- arrestData %>%

group\_by(primary\_type, arrest) %>%

summarize(Total\_Crime = n()) %>%

data.frame() %>%

add\_row(primary\_type = "TOTAL\_CRIME", arrest = FALSE, Total\_Crime = 969746) %>%

add\_row(primary\_type = "TOTAL\_CRIME", arrest = TRUE, Total\_Crime = 193883) %>%

rename("Incidents" = "Total\_Crime") %>%

mutate(primary\_type = recode(primary\_type,

ARSON = 'Arson',

ASSAULT = 'Assault',

BATTERY = 'Battery',

BURGLARY = 'Burglary',

`CONCEALED CARRY LICENSE VIOLATION` = 'Concealed Carry Licence Violation',

`CRIM SEXUAL ASSAULT` = 'Crim Sexual Assault',

`CRIMINAL DAMAGE` = 'Criminal Damage',

`CRIMINAL SEXUAL ASSAULT` = 'Criminal Sexual Assault',

`CRIMINAL TRESPASS` = 'Criminal Trespass',

`DECEPTIVE PRACTICE` = 'Deceptive Practice',

GAMBLING = 'Gambling',

HOMICIDE = 'Homicide',

`HUMAN TRAFFICKING` = 'Human Trafficking',

`INTERFERENCE WITH PUBLIC OFFICER` = 'Interference With Public Officer',

INTIMIDATION = 'Intimidation',

KIDNAPPING = 'Kidnapping',

`LIQUOR LAW VIOLATION` = 'Liquor Law Violation',

`MOTOR VEHICLE THEFT` = 'Vehicle Theft',

NARCOTICS = 'Narcotics',

OBSCENITY = 'Obscenity',

`OFFENSE INVOLVING CHILDREN` = 'Offence Involving Children',

PROSTITUTION = 'Prostitution',

`PUBLIC INDECENCY` = 'Public Indecency',

`PUBLIC PEACE VIOLATION` = 'Public Peace violation',

ROBBERY = 'Robbery',

`SEX OFFENSE` = 'Sex Offence',

STALKING = 'Stalking',

THEFT = 'Theft',

`WEAPONS VIOLATION` = 'Weapons Violation',

TOTAL\_CRIME = 'Total Crime'))

#-------------------------------------------

# Getting glyphs from fontawesomef for isotype

#-------------------------------------------

library(extrafont)

library(showtext)

extrafont::font\_import (path="C:\\Users\\Family\\Downloads", pattern = "fa-", prompt = FALSE)

loadfonts(device = "win")

extrafont::fonttable() %>%

dplyr::as\_tibble() %>%

dplyr::filter(grepl("Awesome", FamilyName)) %>%

select(FamilyName, FontName, fontfile)

font\_add(family = "FontAwesome5Free-Solid", regular = "C:\\Users\\Family\\Downloads\\fa-solid-900.ttf")

font\_add(family = "FontAwesome5Free-Regular", regular = "C:\\Users\\Family\\Downloads\\fa-regular-400.ttf")

font\_add(family = "FontAwesome5Brands-Regular", regular = "C:\\Users\\Family\\Downloads\\fa-brands-400.ttf")

showtext\_auto()

glyph\_font = "FontAwesome5Free-Solid"

# write to working directory for shinydashboard

write.csv(arrestData, "output.csv", row.names=FALSE, quote=FALSE)

***Dashboard code:***

library(shiny)

library(shinydashboard)

data <- merged\_data

data2 <- arrestData

# for waffle/isotype plot

round\_preserve <- function(x, digits = 0) {

up <- 10 ^ digits

x <- x \* up

y <- floor(x)

indices <- tail(order(x-y), round(sum(x)) - sum(y))

y[indices] <- y[indices] + 1

y / up

}

Crime.Variables = c("Total Crime", "Theft", "Battery", "Criminal Damage", "Assault",

"Deceptive Practice", "Vehicle Theft", "Narcotics", "Burglary",

"Robbery", "Weapons Violation")

ui <- dashboardPage(

dashboardHeader(title = "Chicago Crimes Shinydashboard"),

dashboardSidebar(

selectInput(inputId="crime\_variable", label = "Select Crime to Map:",

choices = Crime.Variables, selected = Crime.Variables[1])),

dashboardBody(

tags$head(tags$style(HTML("

#map\_container {

text-align: center;

overflow: auto;

}

")),

tags$script('

// Define function to set height of "map" and "map\_container"

setHeight = function() {

var window\_height = $(window).height();

var header\_height = $(".main-header").height();

var boxHeight = window\_height - header\_height - 90;

$("#map\_container").height(boxHeight);

$("#map").height(boxHeight-20);

};

// Set input$box\_height when the connection is established

$(document).on("shiny:connected", function(event) {

setHeight();

});

// Refresh the box height on every window resize event

$(window).on("resize", function(){

setHeight();

});

')),

fluidRow(

column(6, box(title = uiOutput('map\_title'), solidHeader = TRUE,status="primary",

id = "map\_container", leafletOutput("map"), width=NULL)

),

column(6,

fluidRow(box(title = uiOutput('bar\_title'), solidHeader = TRUE,status="primary",

id = "map\_container", width = NULL,

style = "width: 12; height: 350px; overflow-y: scroll;",

plotlyOutput("plot", height="1300px"))

),

fluidRow(box(title = uiOutput('iso\_title'), solidHeader = TRUE,status="primary",

id = "map\_container", plotOutput("plot2", height = 450),

style = "width: 12; height: 500px;",

width = NULL))

)

)

)

)

server = function(input, output) {

output$map\_title = renderPrint({

HTML(cat("<font size=5>", "<font color='white'> Chicago Reported ", input$crime\_variable,

" By Community Area", "</font>"),

paste("<br>", "<font size=4>", "1/1/2018 - 1/1/2023", "</font>"))

})

output$bar\_title = renderPrint({

HTML(cat("<font size=4.7>", "<font color='white'> ", input$crime\_variable,

" By Community Area", "</font>"),

paste("<br>", "<font size=4>", "1/1/2018 - 1/1/2023", "</font>"))

})

output$iso\_title = renderPrint({

HTML(cat("<font size=4.7>", "<font color='white'> Arrest Rate per 100 Incidents: ",

input$crime\_variable, "</font>"),

paste("<br>", "<font size=4>", "1/1/2018 - 1/1/2023", "</font>"))

})

crimes\_reactive <- reactive({

data %>%

as.data.frame() %>%

group\_by(community, .data[[input$crime\_variable]]) %>%

summarize(total = mean(.data[[input$crime\_variable]], na.rm = TRUE)) %>%

arrange(desc(total))

})

arrests\_reactive <- reactive({

arrest\_summary <- arrestData %>%

filter(primary\_type %in% input$crime\_variable) %>%

summarise(Arrested = round\_preserve(100 \* Incidents[2] / (Incidents[1]+ Incidents[2])),

Not\_Arrested = round\_preserve(100 \* Incidents[1] / (Incidents[1]+ Incidents[2])))

data.frame(

x = rep(1:10, each = 10),

y = rep(1:10, 10),

color = c(rep("blue", arrest\_summary$Arrested), rep("red", arrest\_summary$Not\_Arrested)),

label = fontawesome('fa-male'))

})

output$plot <- renderPlotly({

df\_local <- req(crimes\_reactive())

plot\_ly(df\_local, x = ~total, y=~community, type='bar') %>%

layout(yaxis = list(categoryorder = "total ascending"),

xaxis = list(side ="top", title = 'Reported Incidents'))

})

output$map <- renderLeaflet({

pal <- colorNumeric(palette = "inferno", domain = data[[input$crime\_variable]])

leaflet(data) %>%

addProviderTiles(providers$CartoDB.Positron) %>%

addPolygons(fillColor = ~pal(data[[input$crime\_variable]]),

weight = 2, opacity = 1, color = "lightgrey",

dashArray = "3", fillOpacity = 0.7,

highlight = highlightOptions(

weight = 5, color = "#666", dashArray = "",

fillOpacity = 0.7, bringToFront = TRUE),

label = sprintf("<strong>%s</strong><br/>%g Reports",

data$community, data[[input$crime\_variable]]) %>%

lapply(htmltools::HTML),

group = "community") %>%

addLegend(

pal = pal,

values = data[[input$crime\_variable]],

title = "Reported\nIncidents",

position = "topright") %>%

setView(-87.7, 41.845, zoom = 11)

})

output$plot2 <- renderPlot({

df\_local2 <- req(arrests\_reactive())

ggplot(df\_local2, aes(x, y, colour = color)) +

geom\_text(aes(label=label), family='fontawesome-webfont', size=10) +

coord\_equal() +

theme\_waffle() +

labs(fill = NULL,

colour = NULL) +

theme(plot.title = element\_text(size=15, face="bold"),

legend.text=element\_text(size=13)) +

scale\_color\_manual(labels=c('Arrest', 'No Arrest'), values = c("brown4", "slategray3"),

# increase spacing between legend keys (they were overlapping):

guide = guide\_legend(keyheight = unit(2.5, "lines"), keywidth = unit(2.5, "lines")))

})

}

shinyApp(ui, server)

Mosaic plot code:

library(dplyr)

library(ggplot2)

library(ggmosaic)

library(tidyverse)

library(readxl)

violent <- read\_excel("C:/Users/aleja/OneDrive - DePaul University/violent\_toplocations.xlsx", sheet = "Sheet7")

violent$type <- violent$'Primary Type'

violent$location <- violent$'Location Description'

violent <- violent %>%

mutate(type = factor(type, levels=c("ASSAULT","ROBBERY","CRIMINAL SEXUAL ASSAULT","HOMICIDE"), ordered=TRUE))

str(violent$type)

violent <- violent %>%

mutate(location = factor(location, levels=c("STREET", "APARTMENT", "SIDEWALK", "RESIDENCE", "ALLEY", "PARKING LOT / GARAGE", "SMALL RETAIL STORE", "RESTAURANT", "GAS STATION"), ordered=TRUE))

ggplot(data = violent) +

geom\_mosaic(aes(x = product(location), fill=type)) +

theme(axis.text.x = element\_text(angle = 30, hjust = 1, vjust = 1, size = 12)) +

scale\_fill\_manual(values = c("#BD0026", "#FFCC99", "#CC99FF", "#6699CC")) +

guides (fill = FALSE) +

ggtitle("Most Common Locations of Violent Crime in Chicago") +

theme(plot.title = element\_text(

size = rel(1.5), lineheight = .9)) +

theme(axis.text.y = element\_text(size = 12)) +

ylab("Types of Violent Crimes") +

xlab("Most Common Location Types") +

theme(axis.title.x = element\_text(size = 12)) +

theme(axis.title.y = element\_text(size = 12))

HEAT MAP CODE:

