

When and where do most robberies occur in the city of Chicago?

Sooihk Ro

3/5/2021

ABSTRACT

The city of Chicago has a reputation as America's highest crime ridden city. Having experienced robberies, I was interested in exploring the crime dataset provided by the Chicago Police Department to understand when and where robberies occur most often in the city. This exploration can help us become aware of how to safely navigate through the city.

Question

The aim of this project is to answer the following question: At what time of day and locations do most robberies occur in the city of Chicago?

Data and Methods

The Crimes - 2001 to Present dataset provided by the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) will be analyzed to answer the question (Link: <https://data.cityofchicago.org/api/views/ijzp-q8t2/rows.csv?accessType=DOWNLOAD>). The dataset reflects reported incidents of crimes that took place in the city of Chicago from 2001 to present, not included is the most recent seven days and is updated daily. As of February 26, 2021, the dataset contains over 7.28 million observations(rows) and 22 variables(columns). In order to use the data to answer the question, variables in the set were created, transformed, and cleaned of missing values. Visual analysis was performed by creating barplots, lineplots and heatmaps. Rstudio (IDE for R, a programming language for statistical computing and graphics) was used to create the mentioned plots while Tableau (business intelligence visualization tool) was utilized to visualize robbery hotspots on a map and to create tables. Rmarkdown was used to write this report.

Preparation of the dataset

These are the libraries used in RStudio for this report:

```
library(ggplot2)
library(lubridate)
library(zoo)
library(dplyr)
library(scales)
```

```
library(knitr)
library(png)
```

Load the crime dataset into R and show the 22 columns.

These are columns in the dataset:

```
## [1] "ID" "Case.Number" "Date"
## [4] "Block" "IUCR" "Primary.Type"
## [7] "Description" "Location.Description" "Arrest"
## [10] "Domestic" "Beat" "District"
## [13] "Ward" "Community.Area" "FBI.Code"
## [16] "X.Coordinate" "Y.Coordinate" "Year"
## [19] "Updated.On" "Latitude" "Longitude"
## [22] "Location"
```

The original dataset was then filtered to only include robberies. The date and time of the crime was separated from the Date variable in order for R to understand the value. Hour, Count and weekday variables were created to hold more detailed information. Robberies in 2021 and 2001 were omitted since 2021 is not over and 2001 had too many missing values. Missing data was then omitted from the dataset. A variable called locationCat was created to categorize all locations of robberies from Location.Description variable to 9 distinct values. The following is the code used to transform the dataset to answer the question of this paper:

```
# obtain only relevant variables that pertain to answering the question
# narrow dataset to only include robberies
new_crimeData <- select(crime_rawData, Date, Year, Primary.Type, Location.Des
cription,
                        District, Community.Area, Latitude, Longitude) %>%
  filter(Primary.Type == "ROBBERY")

# Make transformations on Date variable in order for R to understand
new_crimeData <- mutate(new_crimeData, Date = mdy_hms(Date))
# Make new variable Hour and extract hour from Date
new_crimeData <- mutate(new_crimeData, Hour = substring(Date, 12, 13))
new_crimeData$Hour <- as.numeric(new_crimeData$Hour)
# Make new count variable, needed for aggregations
new_crimeData$Count <- 1
# Remove time from Date variable
new_crimeData$Date <- as.Date(new_crimeData$Date, format = "%m/%d/%Y")
# Create another variable named weekday to obtain day of the robbery
new_crimeData$weekday <- weekdays(new_crimeData$Date)
# remove 2021 year since it is not complete
new_crimeData <- new_crimeData[!new_crimeData$Year > 2020, ]
# remove 2001 year since so many variables have missing values
new_crimeData <- new_crimeData[!new_crimeData$Year < 2002, ]
# omit rows with missing data
new_crimeData <- na.omit(new_crimeData)
```

The transformed dataset has now 248,467 observations composed into 12 variables. These are the new columns:

```
colnames(new_crimeData)
```

```
## [1] "Date"          "Year"          "Primary.Type"
## [4] "Location.Description" "District"      "Community.Area"
## [7] "Latitude"      "Longitude"     "Hour"
## [10] "Count"        "weekday"       "locationCat"
```

Visualize and Explain Data

The trend of robbery in Chicago

We must first get an overall perspective of how robbery in Chicago evolve over time. Figure 1 shows that daily robberies has been decreasing throughout the years with the exception of two bumps increasing then decreasing. Figure 2 shows the sum of robberies per year. The plot highlights that robberies in 2008-2009 and 2016-2017 spiked, but the overall trend is that robberies in Chicago trends downwards. The sum of robberies in 2002 is lower than the following years is because the date contained many missing values which resulted in removal of many observations. The next section goes further into detail of when most robberies occur.

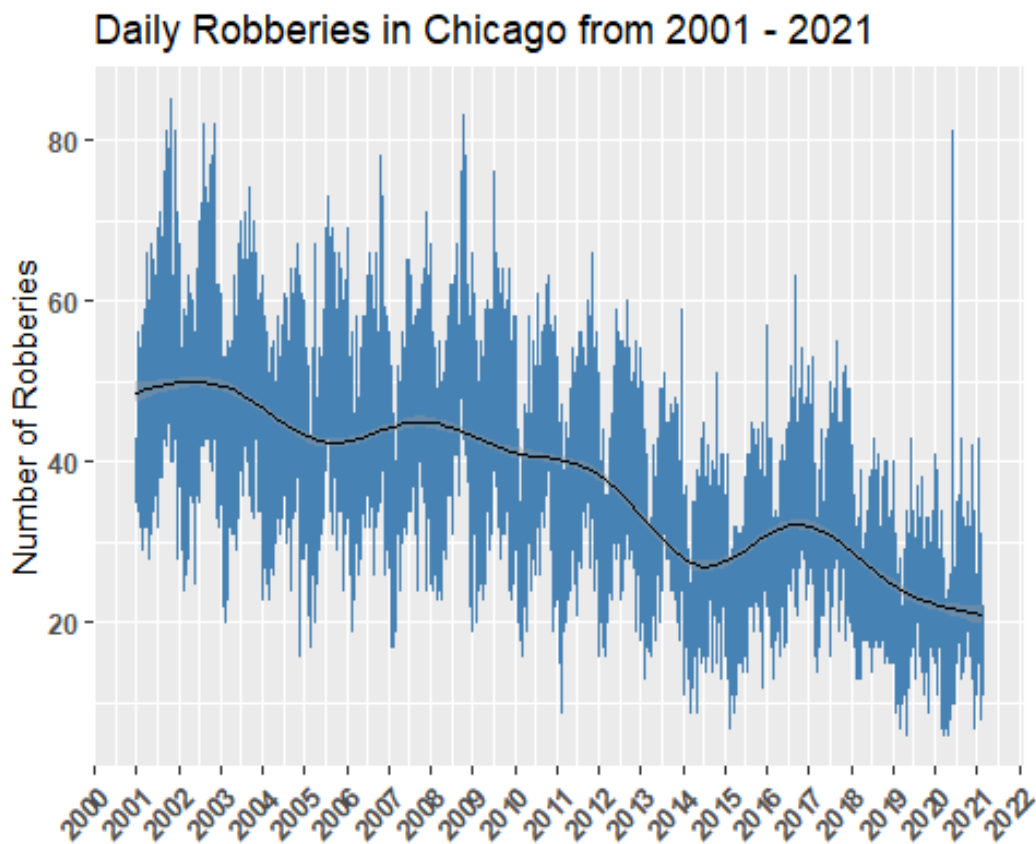


Figure 1

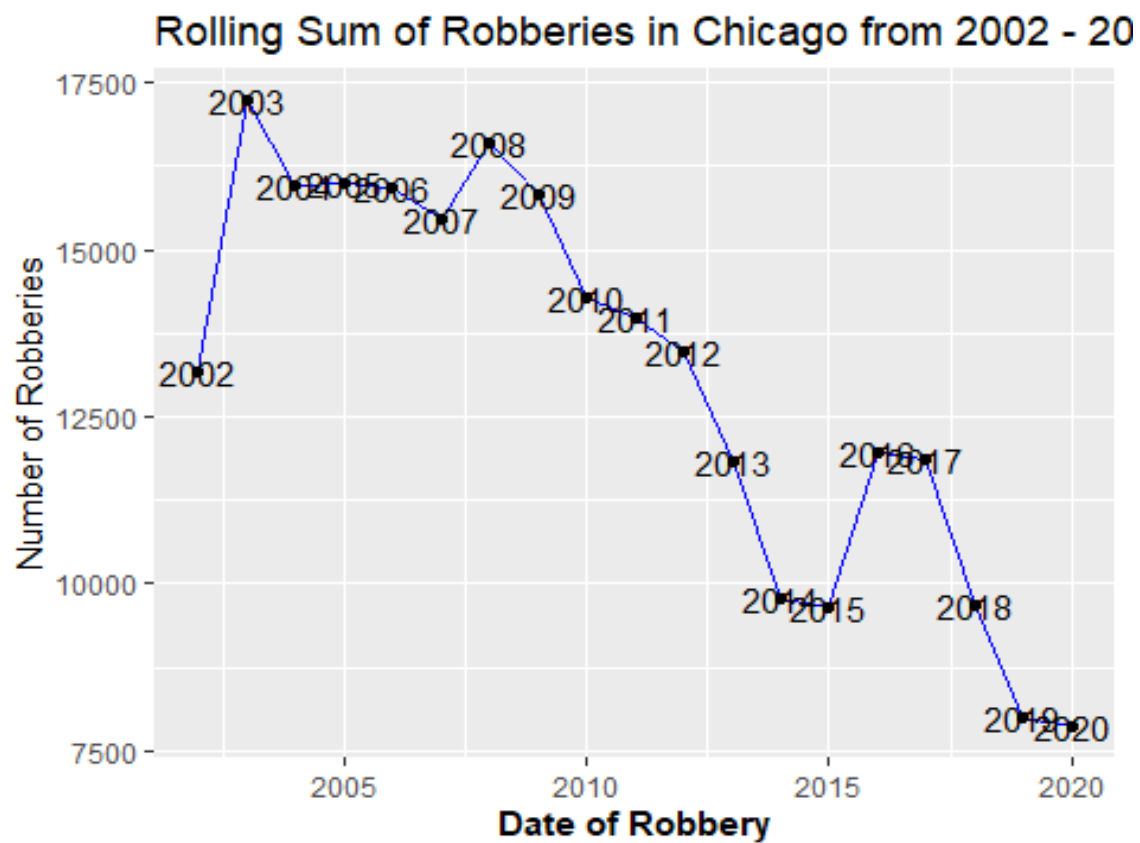


Figure 2

When do most robberies occur?

As we check quarterly robberies from 2001 to 2021 in Figure 3, we see a periodic trend where most robberies occur in quarter 3 and 4 while the least amount occur in quarter 1 of each year. Quarter 3 (July-September) is during good weather and quarter 4 (October-December) is during festive holidays when more people are out, giving more chances for robberies. Quarter 1 is during the winter (January-March) with the cold temperatures ranging on average below freezing to 20 F. Less people are outside during this cold period and presents less opportunities for robberies to occur, correlating to the lowest number of robberies per year.

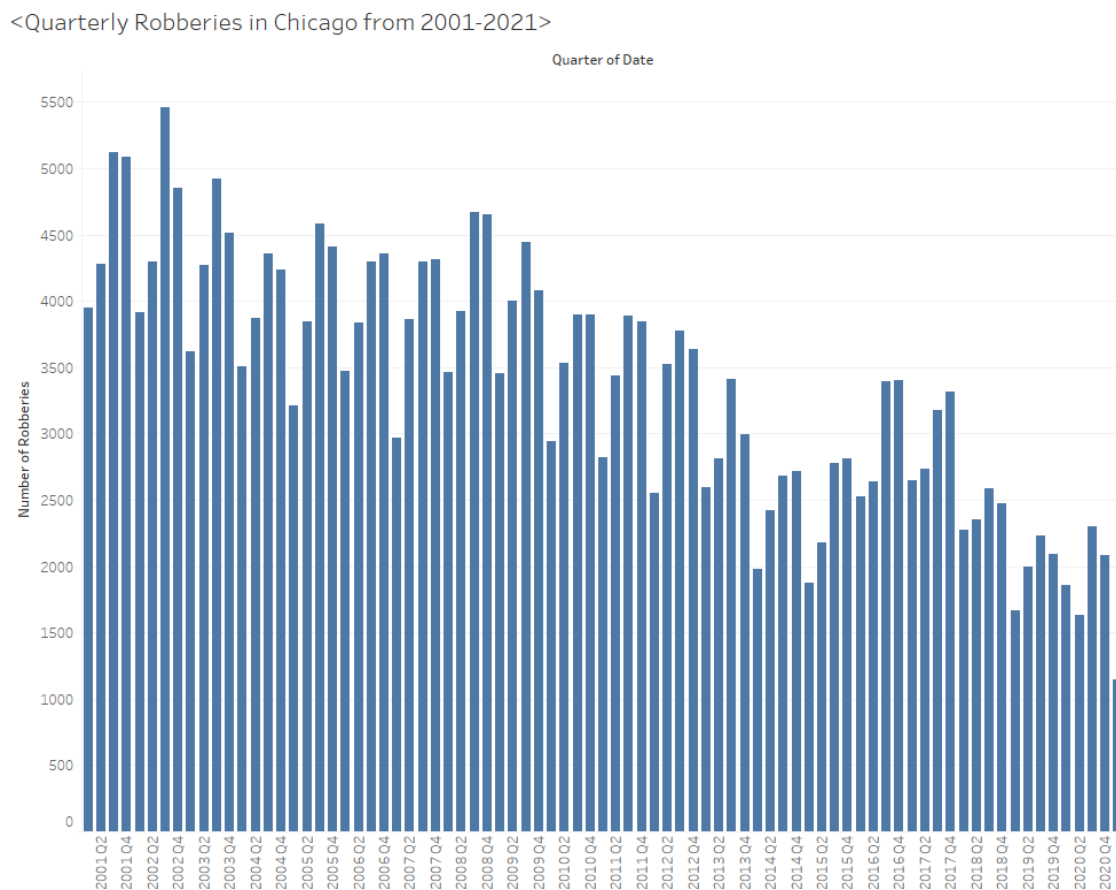


Figure 3

The cross table (Figure 4) made in Tableau further explores which months have the tendency to have the highest occurrence of robberies. The table shows the total count of robberies from 2001-2021 by Month vs Day. The table represents February and March having the lowest occurrence of robberies (Yellow) compared to the months of July to November (Dark Green). February 29th has the lowest record of robberies by day shown as red because it is a leap year.

<Total Robberies Month vs Day 2001-2021>

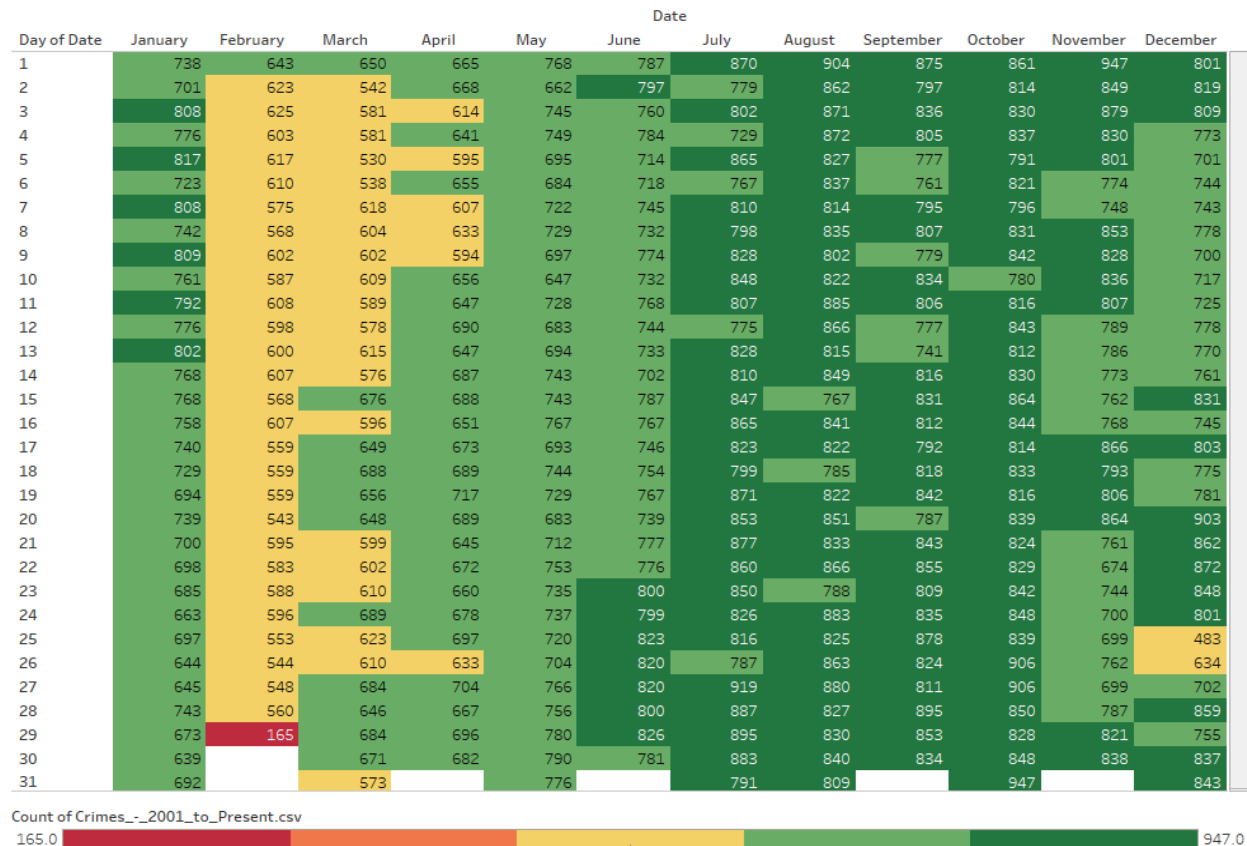


Figure 4

Figure 5 delves further into detail of what time of day most people are robbed. The bar graph shows a gradual increase of robberies from 7am (the least occurrence of robberies) till 9pm (hour 21 with the highest number of robberies). The highest number of robberies takes place during the hours of 18 to 23 (6pm to 11pm). This can be explained as during this time it is darker, and more people are out in the street doing activities.

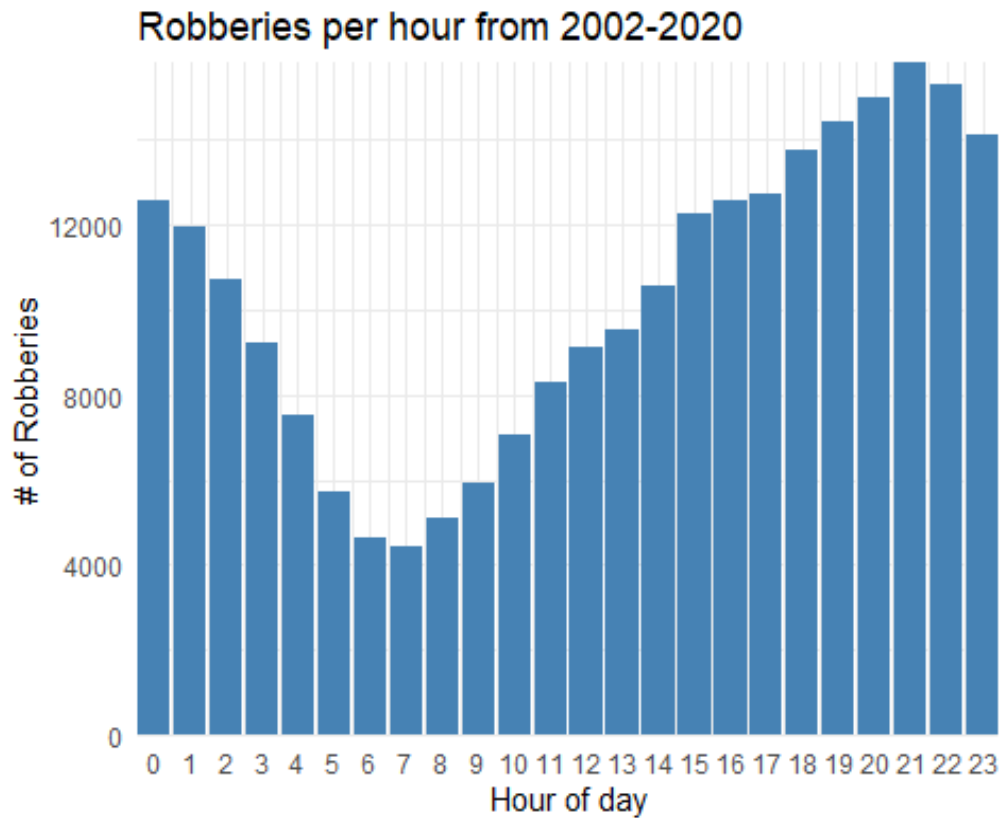


Figure 5

A heatmap was created to be visualize which day of the week and hour of day when most robberies occur, as depicted in Figure 6. The heatmap shows the distribution of the amount of robberies per hour per day with blue being the lower end of the specturm while red the higher end. The heatmap shows that during the weekdays (Monday-Friday), most robberies occur from 3pm to 11pm with 9pm having the most robberies. Friday at 9pm and 10pm is especially concentrated with robberies compared to rest of the week. Saturday and Sunday robberies cluster later in the day around 9pm to 3am with the most robberies occurring at 2am.

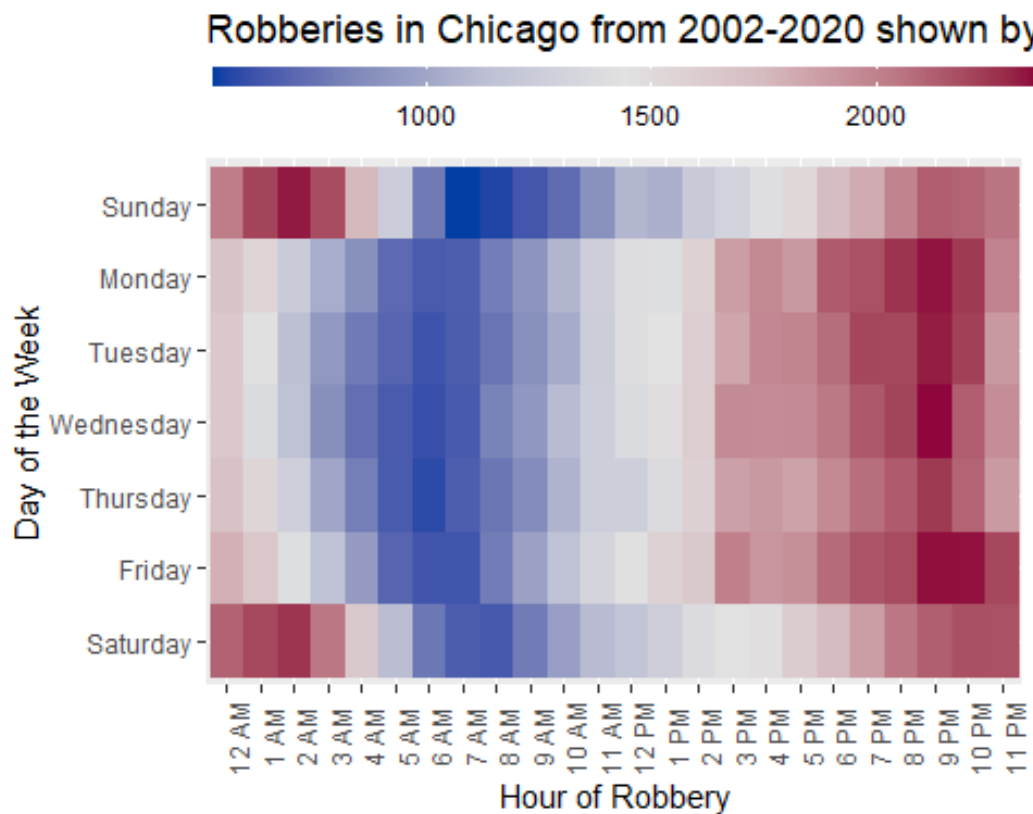


Figure 6

Where do most robberies occur?

Figure 7 shows that the vast majority of people get robbed on the street followed by at small businesses, residential areas and “others”.

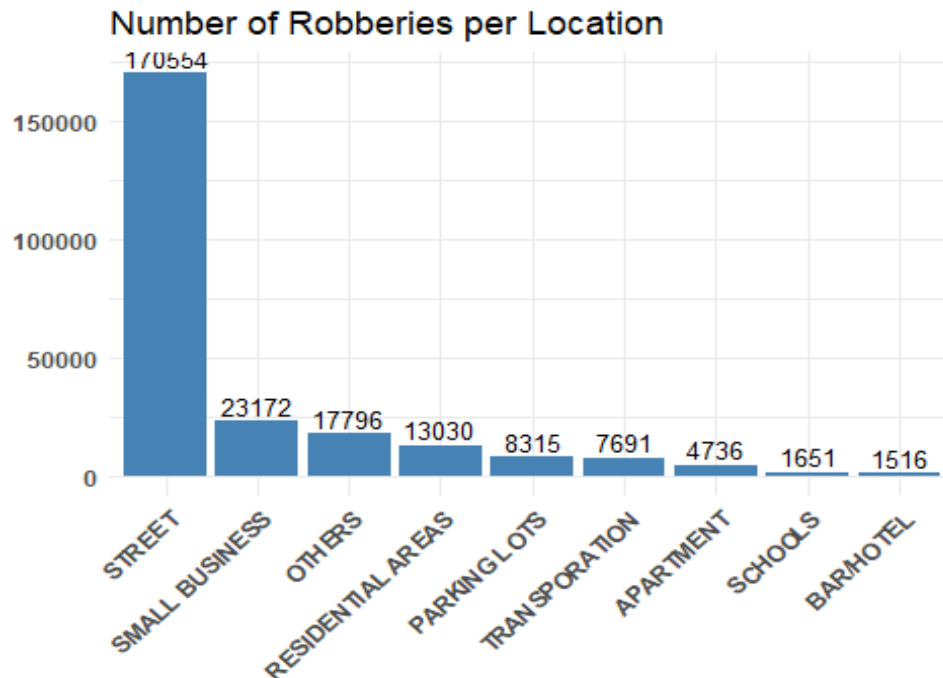


Figure 7

Another barplot was created (Figure 8) to illustrate which community areas have the highest incidences of robberies. Community area 25 (Austin) had the highest occurrence of robberies of up to 17,911 reports since 2002 followed up area 43 (South Shore). Community areas 71, 69, 29, 23, 66, 68, 67, 44, 49 and 24 share similar number of robberies ranging near 7600-8600.

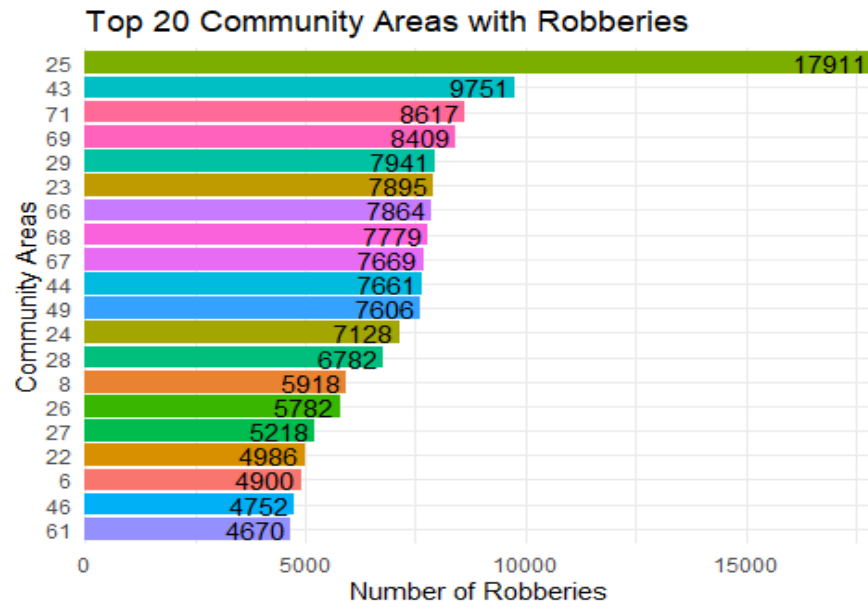


Figure 8

Figure 9 is a community area map of Chicago.

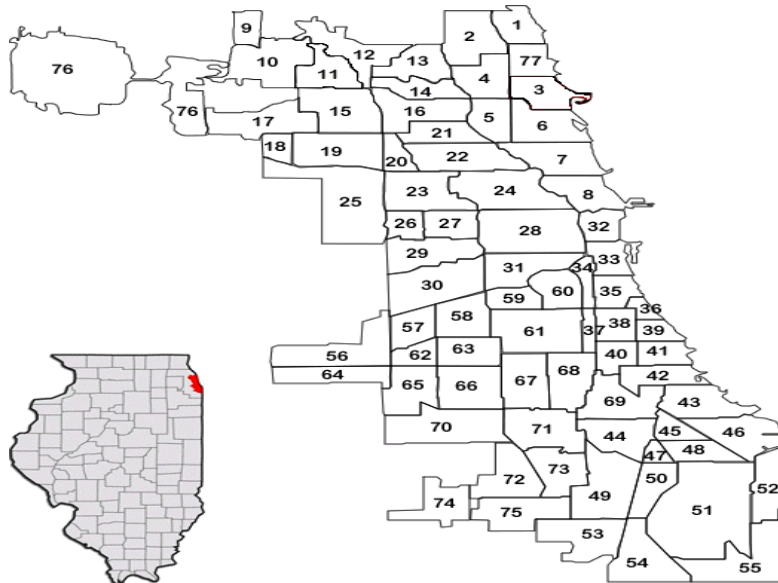


Figure 9

Figure 10 and 11 made on Tableau visualize the areas of where robberies occur. Figure 10 shows the sum of robberies from 2001-2021 per community area illustrating the areas with more robberies as larger circles. Figure 11 is a heat intensity map using data from 2020. In Figure 11, the heat intensity map visualizes brighter areas of the map for high incidents of robberies while the darker areas represent less incidents of robberies. Other years show a similar heat intensity map as 2020 but only one year was chosen to create this map to avoid visual clutter. Both figures show 3 clusters where robberies occur the most. The zone with the highest amount of robberies is community area 25 and the 7 areas east of area 25. The second zone with robberies more spread out than the previous zone is the region south of community area 61 and north of area 49. The third zone with high numbers of robberies are the community areas located at edge of the coast northeast of Chicago (The line of areas south of area 1 and north of area 33).

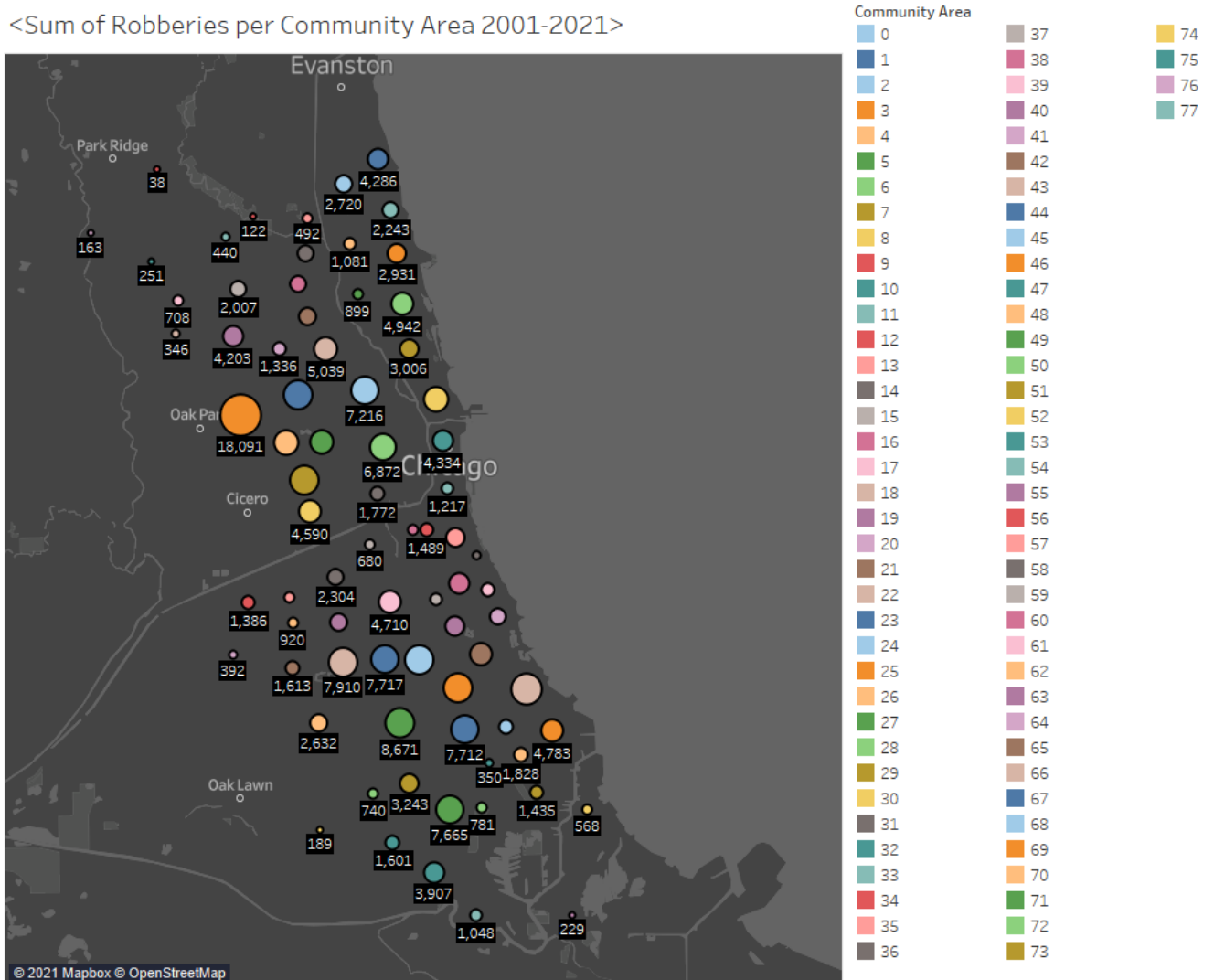


Figure 10

<Heat Intensity Map of Robberies in 2020>

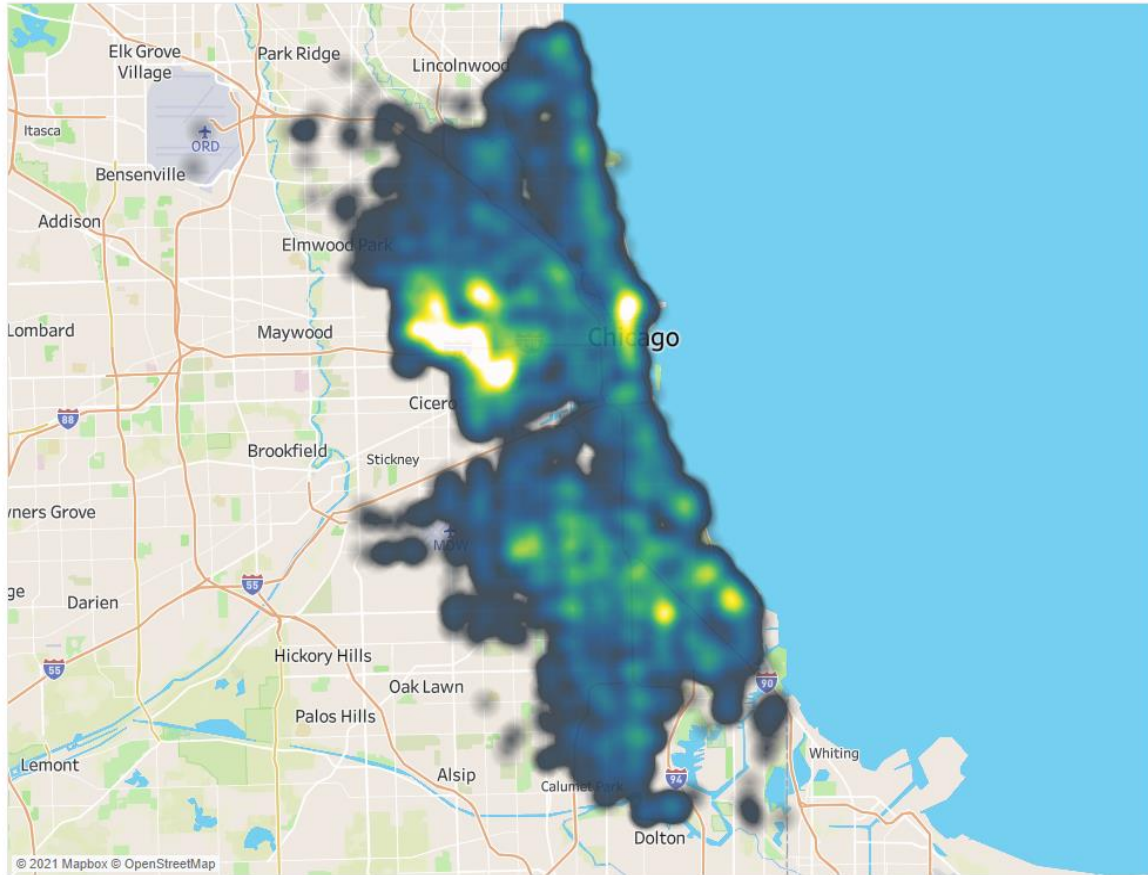


Figure 11

Conclusion

RStudio and Tableau were the tools used to create plots to visually answer when and where most robberies occur in the city of Chicago. Most robberies occur during the months of July to October every year. On weekdays, people are usually robbed between 6pm to 10pm, peaking at 9pm. On weekends, robberies mostly occur between 9pm to 3am, peaking at 2am. Vast majority of robberies occur on the street and in community area 25 (Austin) with 17,911 reports followed by area 43 (South Shore) with 9751 reports. There are 3 main clusters where robberies occur the most often. One is the zone of area 25 and east of area 25, and the other cluster is the zone of community areas surrounding area 69. The last cluster is the strip of coastal community areas south of area 1 and north of area 32.

Figures

Plots Made in RStudio

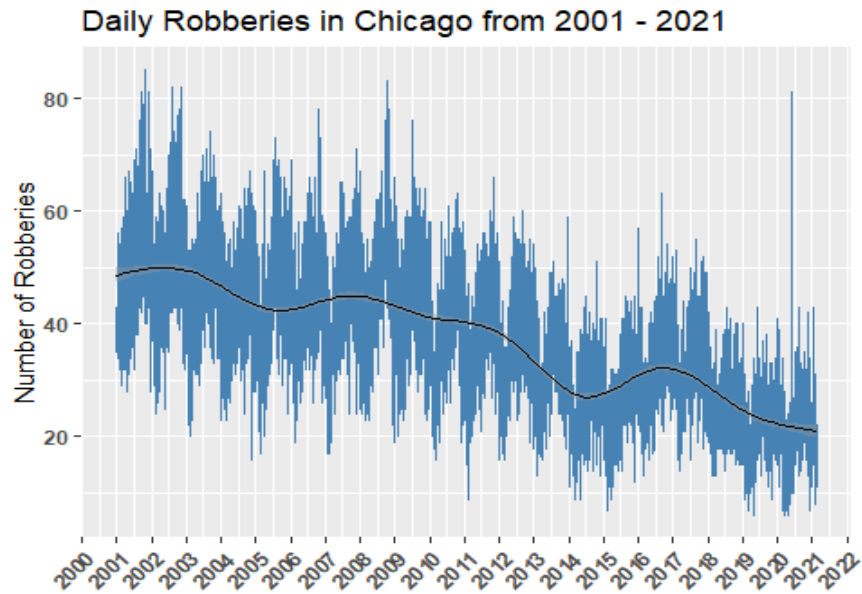


Figure 1

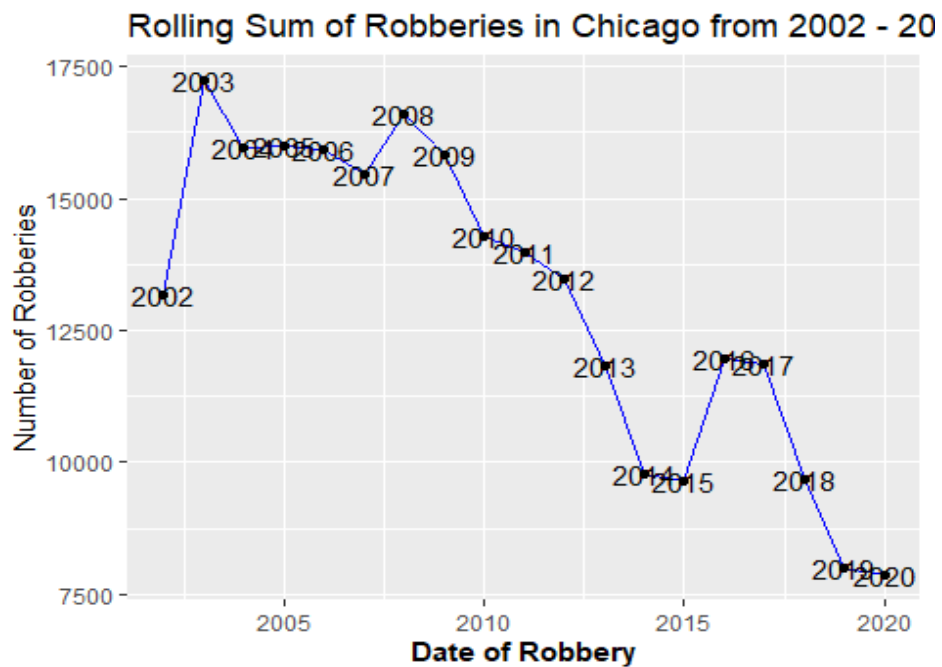


Figure 2

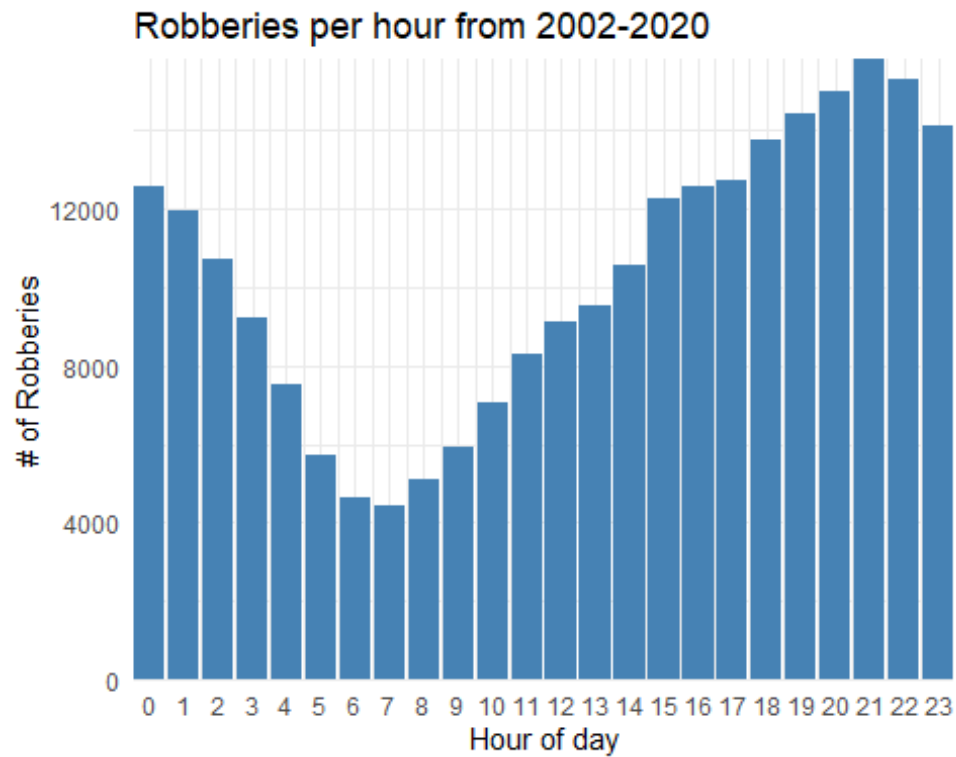


Figure 5

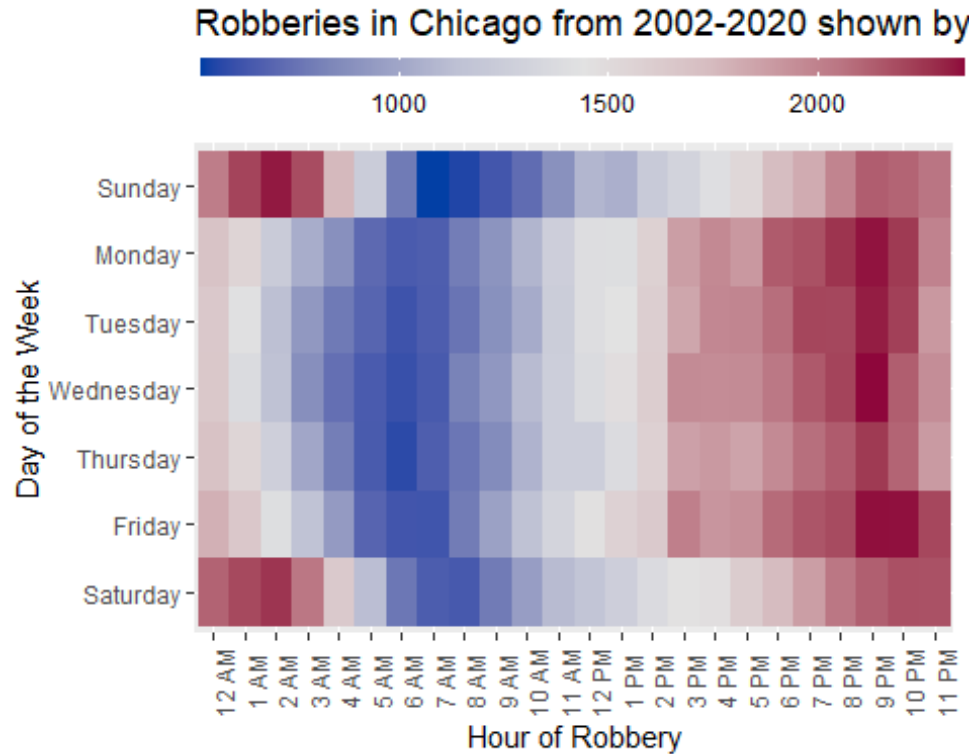


Figure 6

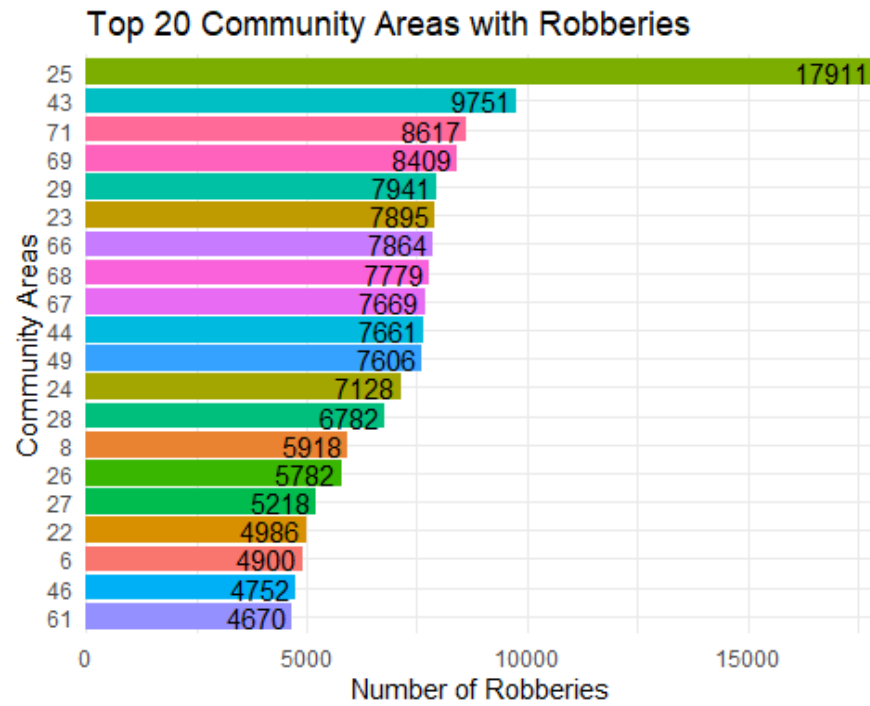


Figure 8

Figures Made in Tableau

<Sum of Robberies per Community Area 2001-2021>

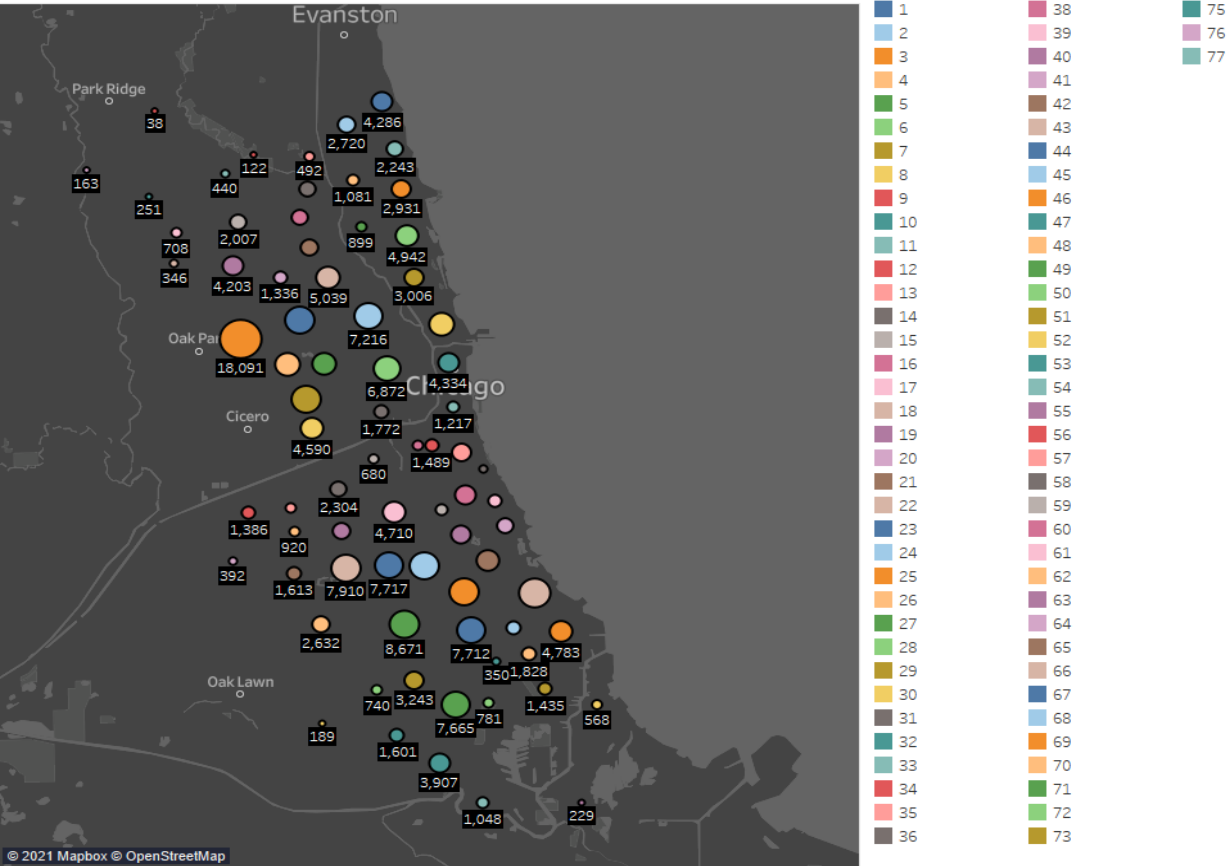


Figure 10

<Heat Intensity Map of Robberies in 2020>

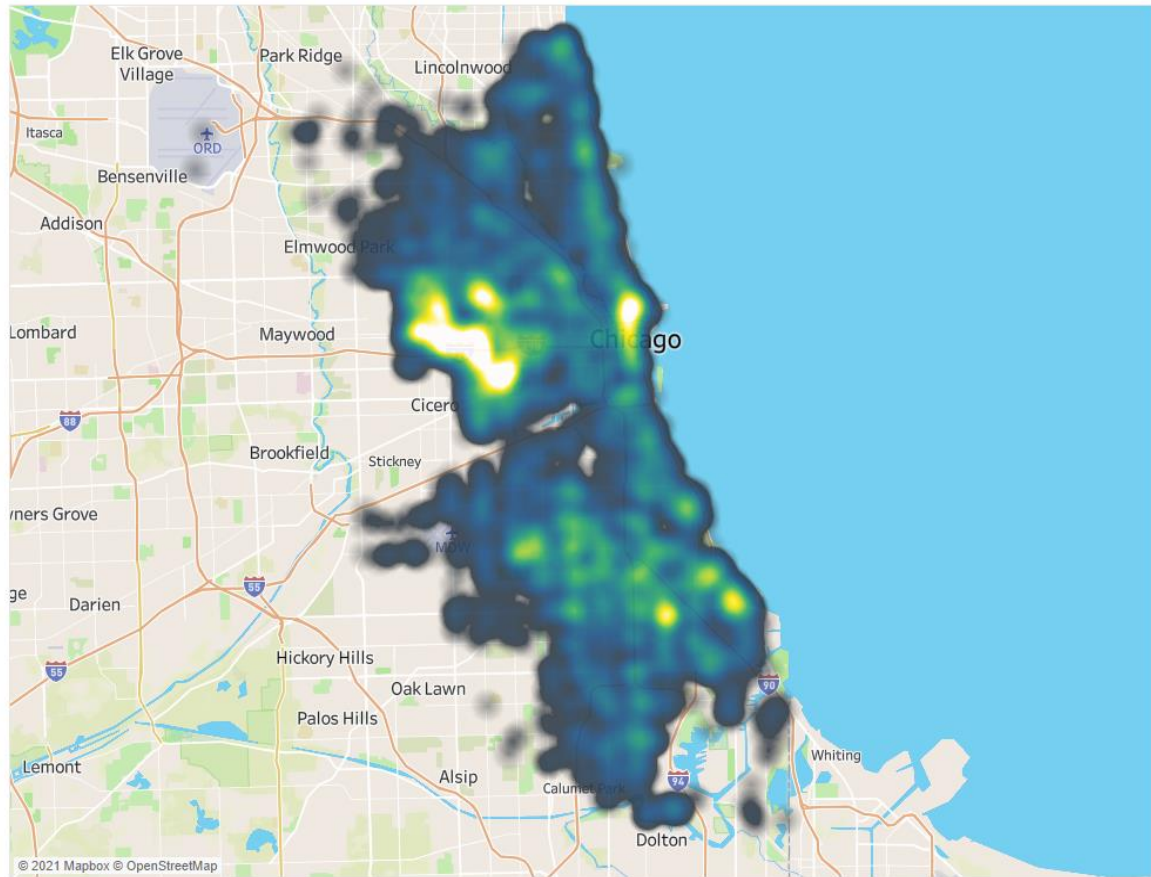


Figure 11

<Quarterly Robberies in Chicago from 2001-2021>

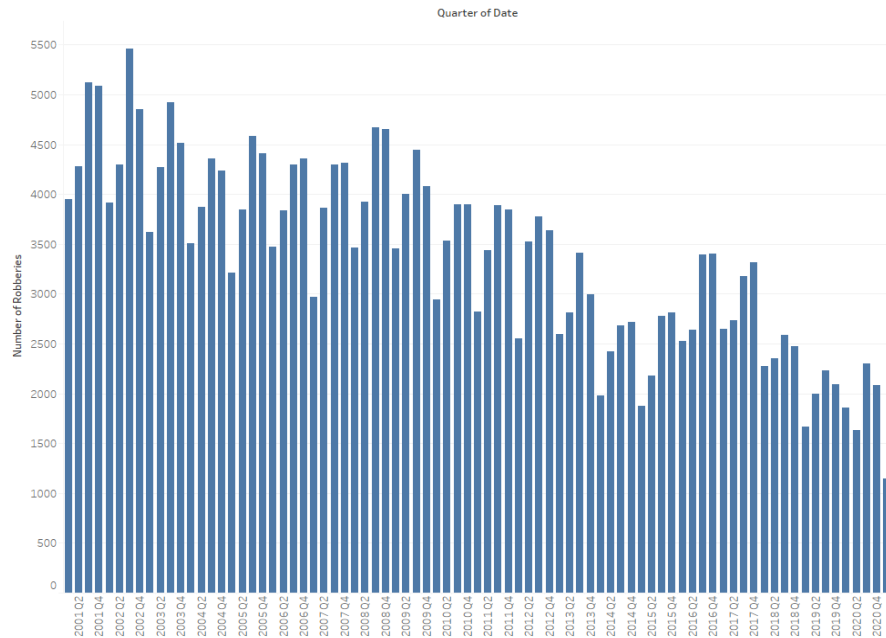


Figure 3

<Total Robberies Month vs Day 2001-2021>

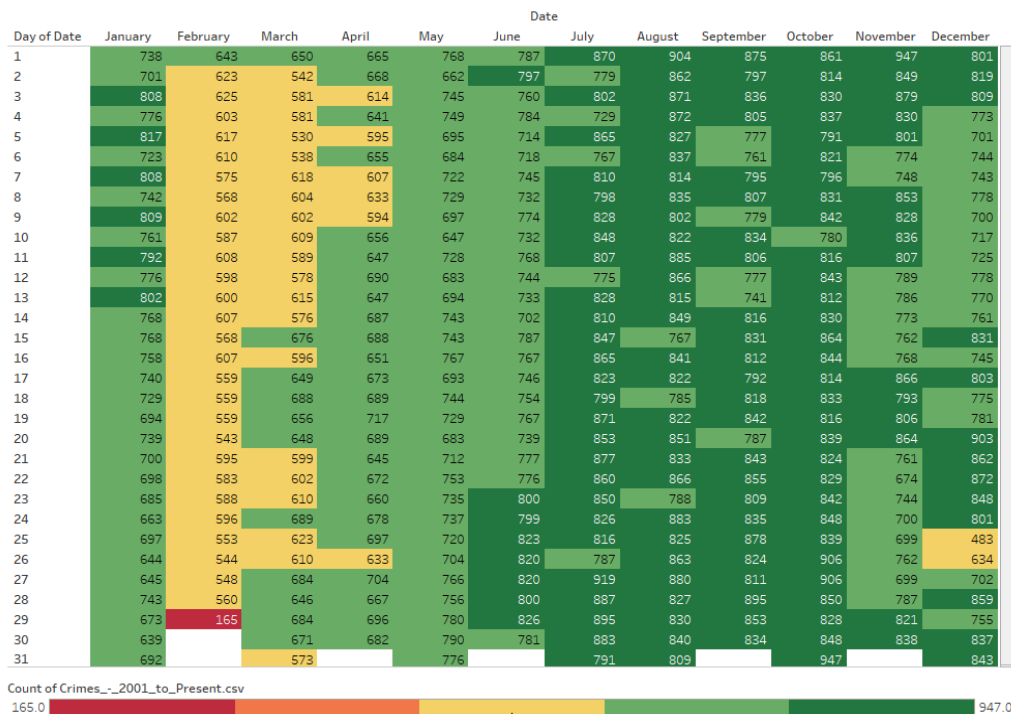


Figure 4