**Objective**

My objective was to examine the top 5 crimes based **only** on number of “reported” occurrences since 2001. Specifically, I divided the entire time frame into four (4) time intervals as follows:

* + 2001 – 2005
  + 2006 – 2010
  + 2011 – 2015
  + 2016 to present

A total of 6,900,557 crimes were reported for the entire period from 2001 – present. I examined the trends in these crimes with time as well as seasonal variation (months of the year) during each time interval.

The top 5 crimes based on reports are:

|  |  |  |
| --- | --- | --- |
| **Crime** | **Occurrences during the Entire Timeframe** | **Percent of Total** |
| Theft | 1,453,930 | 21.1 |
| Battery | 1,260,969 | 18.3 |
| Criminal Damage | 786,779 | 11.4 |
| Narcotics | 719,644 | 10.4 |
| Assault | 430,611 | 6.2 |
| **TOTAL** | **4,651,933** | **67.4** |

Thus, I have examined over 65 percent of the crimes reported for the entire period 2001-present.

**Data and Data Source**

I downloaded the raw data, “Crimes\_-\_2001\_to\_present.csv” from the following location:

[www.data.gov](http://www.data.gov)

This is a fairly comprehensive data file containing more than 7 million records.

**Data Selection and Data Cleaning**

My task was to examine the top five (5) crime “types” purely based on the number of occurrences. It is true that some crimes, even though reported to be a small percent of the almost 7 million reports, may be much more severe in terms of their consequences. Emotional considerations were set aside during this evaluation.

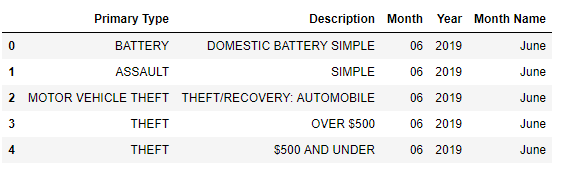
The file had the following fields in that order from left to right:

|  |  |
| --- | --- |
| * + ID   + Case Number   + Date   + Block   + IUCR   + Primary Type   + Description   + Location Description   + Arrest   + Domestic   + Beat   + District   + Ward   + Community Area   + FBI Code | * + X Coordinate   + Y Coordinate   + Year   + Updated On   + Latitude   + Longitude   + Location   + Historical Wards 2003-2015   + Zip Codes   + Community Areas   + Census Tracts   + Wards   + Boundaries – Zip Codes   + Police Districts   + Police Beats |

As a first step, I picked only the columns that were relevant for my part of the study as follows:

* + Primary Type
  + Description
  + Date
  + Year

I then extracted the Month number from the date field and defined two new columns, “Month” and “Month Name” and defined a separate dataframe.[[1]](#footnote-1) A screenshot of the header of this dataframe is shown below:



I then cleaned the dataframe using the “dropna(how = ‘any’)” method. At the end of this step, I had a total of 6,900,774 records. I then deleted Rows which had “Non – Criminal” as the “Primary Type” which gave me a total of 6,900,557 Rows.

This means I have “valid” data for the selected columns and that further analysis can be initiated.

**Determination of Time Intervals**

I divided the data to represent four (4) time intervals as follows:

|  |  |
| --- | --- |
| * + 2001 – 2005   + 2006 – 2010 | * + 2011 – 2015   + 2016 to present |

Accordingly, I developed four dataframes based on performing a “df.loc” on the column “Month”. Even though this may have involved “more than necessary” code, I took this step to keep things clean and separate in the event any additional analyses are carried out for specific time intervals based on interesting observations and trends.

For each time interval, I obtained individual “monthly” counts of for each one of the time periods described above:

* + - Theft
    - Battery
    - Criminal Damage
    - Narcotics
    - Assault

I defined dataframes to hold these values and plotted them as bar charts.

**Results and Inferences**

Appendix A presents raw crime data for each of the five-year intervals. However, the Tabular form is not user-friendly. Appendix B presents bar charts for the same crimes whose data are presented in Appendix A. The bar charts are much more user-friendly. Finally, the trend for each one of these crimes seems to be dropping as shown in Appendix C.

Following inferences were made based on the analysis:

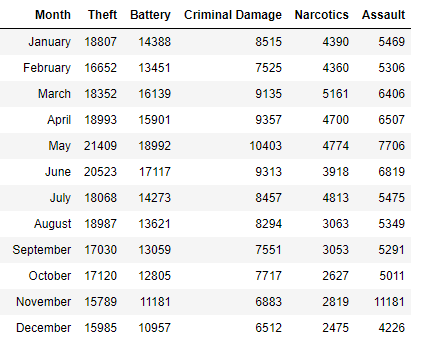
* Most of the selected crimes take place during the time when there are clement weather conditions (May through August)
* However, December and January show a much higher level of “thefts” compared to the other slow months
* The five crimes have been showing a decreasing trend since 2001 which is the first year of the dataset

**APPENDIX B**

**CRIME DATA FOR FIVE-YEAR INTERVALS**

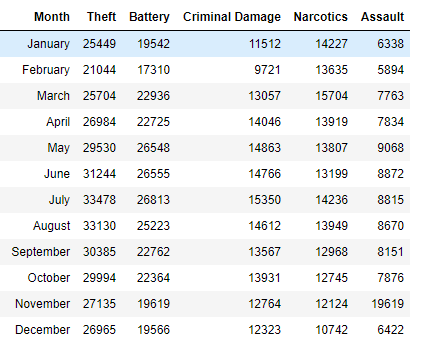
**TABLE A-1**

**CRIME STATISTICS BY MONTH FOR 2016 AND LATER**



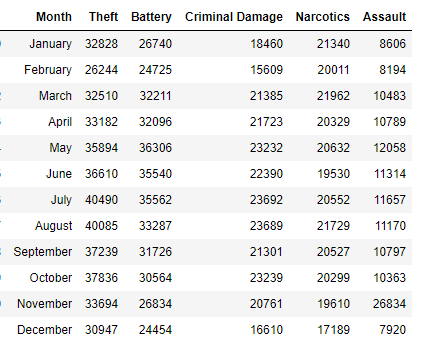
**TABLE A-2**

**CRIME STATISTICS BY MONTH FOR 2010 TO 2015**



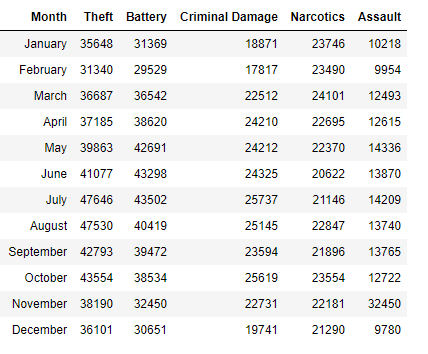
**TABLE A-3**

**CRIME STATISTICS BY MONTH FOR 2006 TO 2010**



**TABLE A-4**

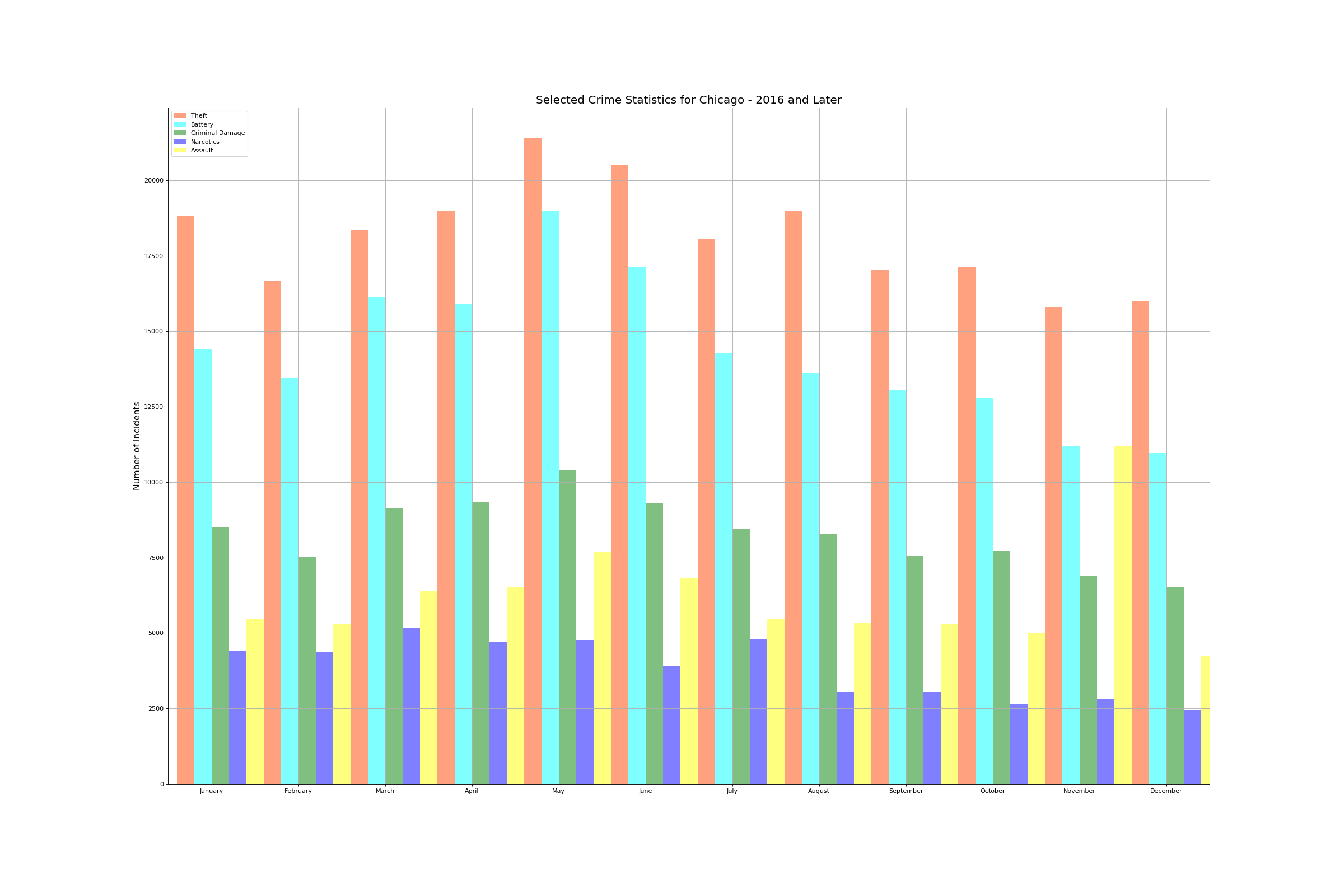
**CRIME STATISTICS BY MONTH FOR 2001 TO 2005**



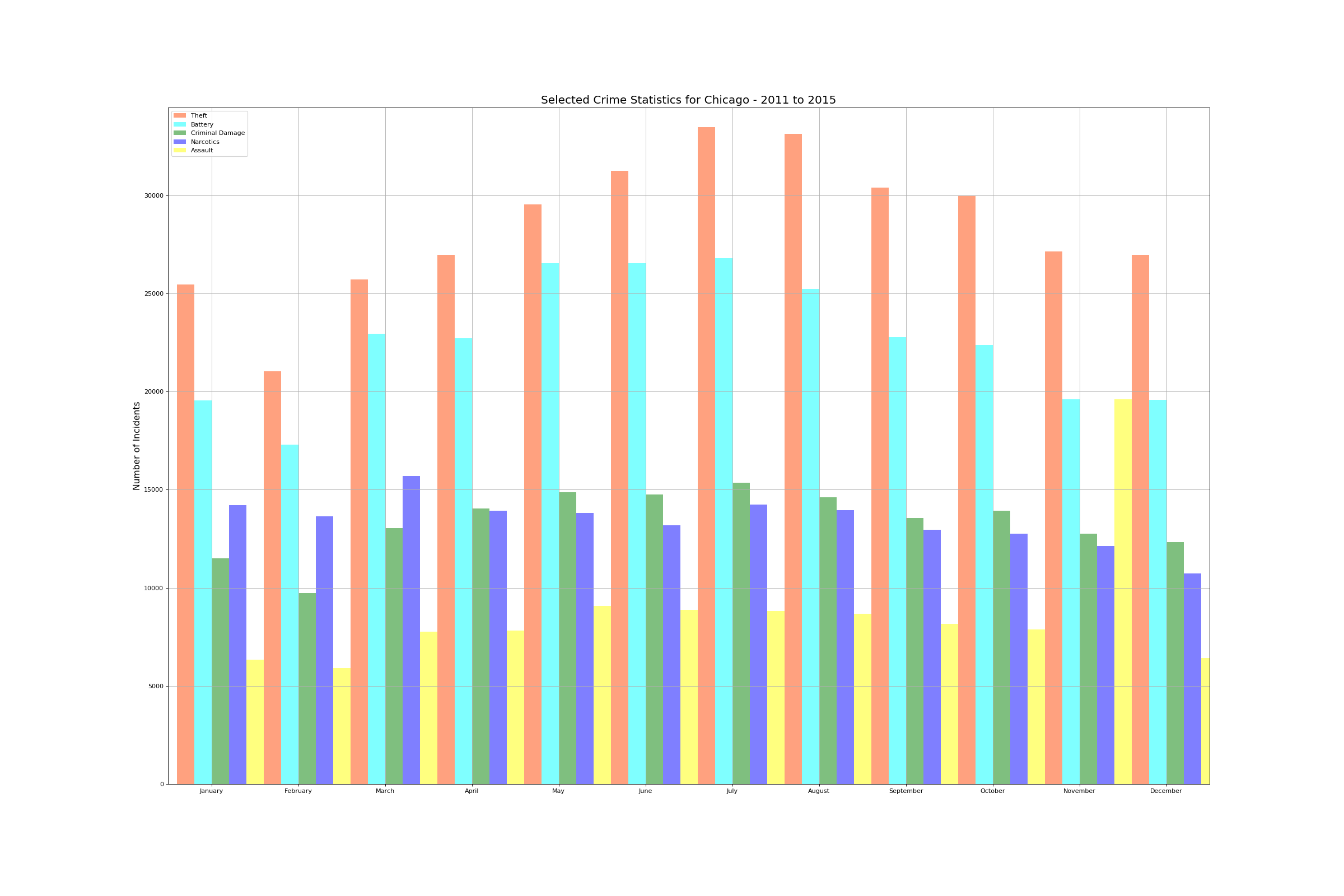
**APPENDIX B**

**BAR CHARTS**

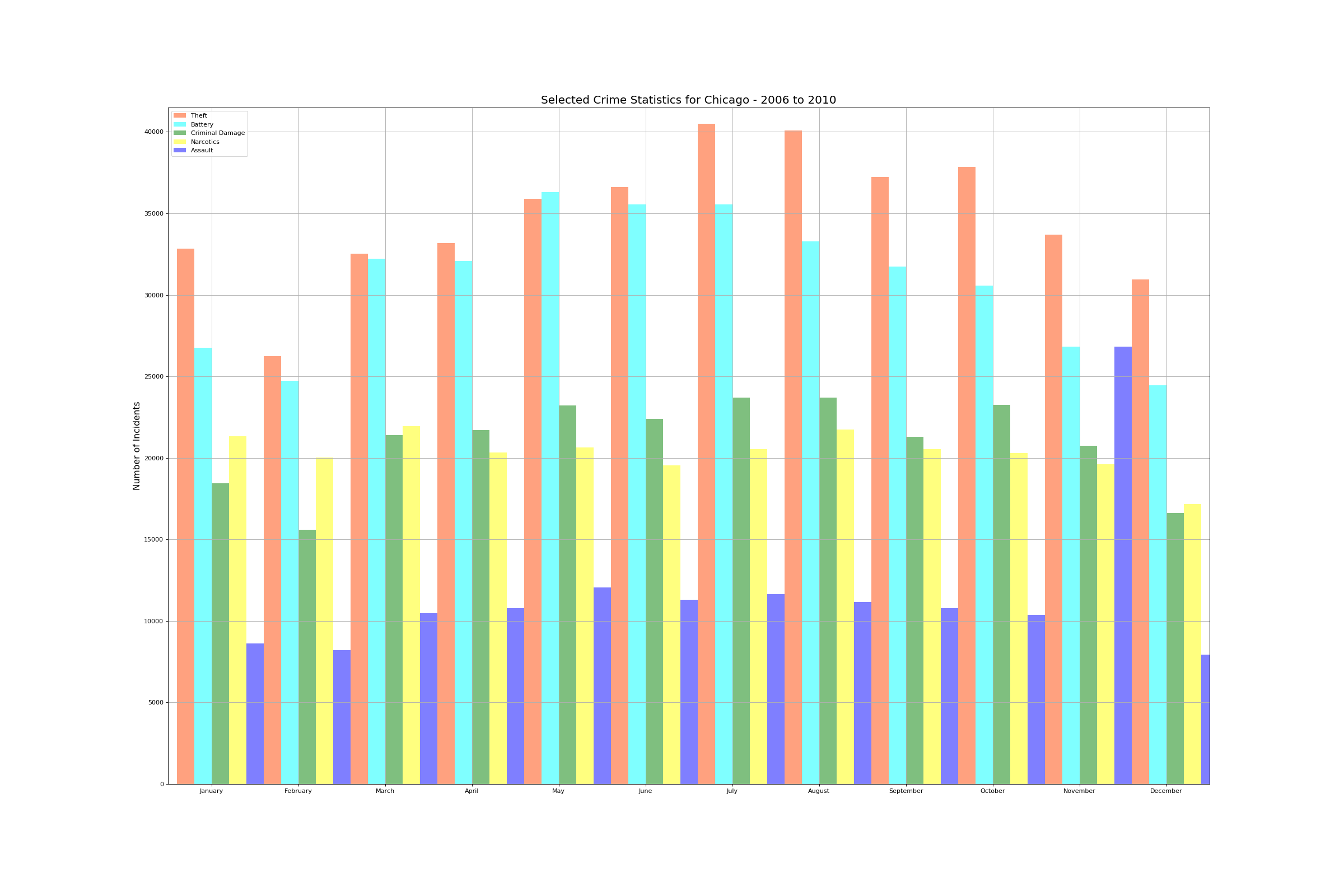
**FIGURE B-1**



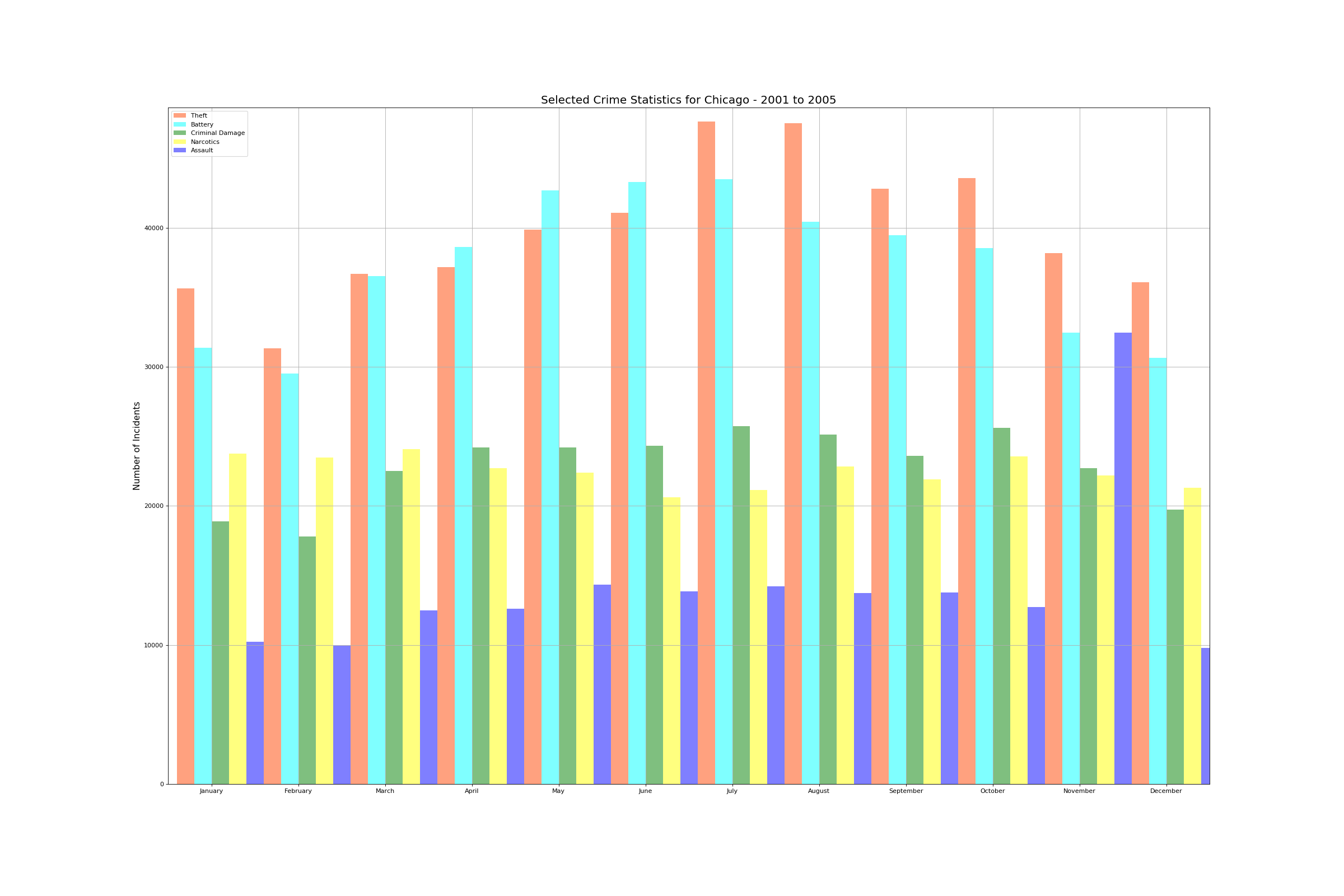
**FIGURE B-2**



**FIGURE B-3**



**FIGURE B-4**

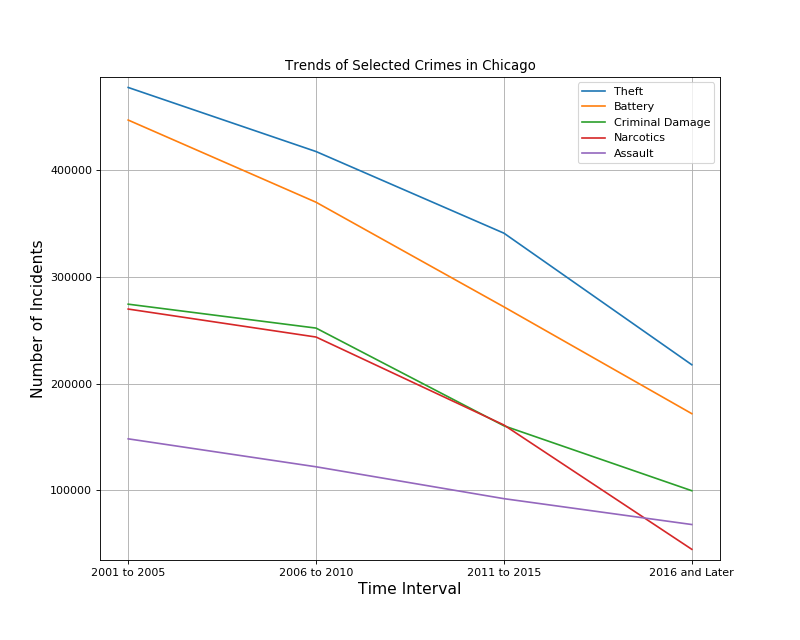


**APPENDIX C**

**TREND OF SELECTED CRIMES IN CHICAGO – 2001 TO PRESENT**

**FIGURE C-1**

**TREND OF SELECTED CRIMES IN CHICAGO (2001 TILL PRESENT)**



1. I wrote a piece of code which looked at the two digits in the “Date” field and assigned the month name based on if conditions. [↑](#footnote-ref-1)