



Crime and the CTA

Team Git Pull Origin Laptop: Emma, Gregory,
Sushank and Emmarose



Agenda

- Research questions and hypothesis
- Data sources and cleaning
- Data analysis
- Discussion
- Further steps
- Q&A



Research Questions

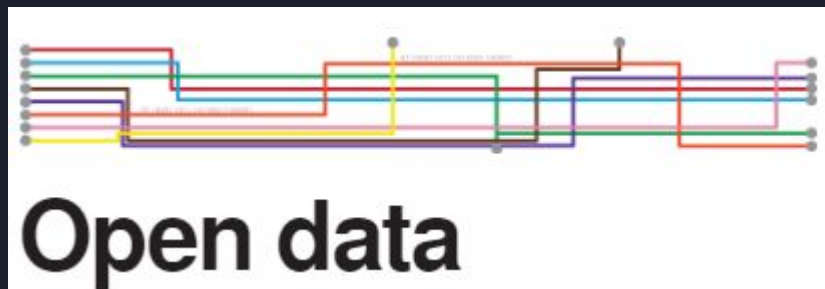
Hypothesis: If ridership on CTA trains increases, then crime will increase as well.

Research questions:

- What is the relationship between high ridership and crime?
- How does ridership change after a crime happens?
- What types of crime are highest on CTA?
- How do special events (ex. Holidays, events) impact crime and ridership?

Data: CTA

| | Date | Day Type | Bus | Rail Boardings | Total Rides |
|------|------------|----------|--------|----------------|-------------|
| 0 | 2001-01-01 | U | 297192 | 126455 | 423647 |
| 1 | 2001-01-02 | W | 780827 | 501952 | 1282779 |
| 2 | 2001-01-03 | W | 824923 | 536432 | 1361355 |
| 3 | 2001-01-04 | W | 870021 | 550011 | 1420032 |
| 4 | 2001-01-05 | W | 890426 | 557917 | 1448343 |
| ... | ... | ... | ... | ... | ... |
| 6996 | 2019-12-27 | W | 552198 | 445835 | 998033 |
| 6997 | 2019-12-28 | A | 394869 | 298646 | 693515 |
| 6998 | 2019-12-29 | U | 315428 | 251105 | 566533 |
| 6999 | 2019-12-30 | W | 544713 | 480721 | 1025434 |
| 7000 | 2019-12-31 | W | 511289 | 431677 | 942966 |



Data: Crime



| | ID | Case Number | Date | Block | IUCR | Primary Type | Description | Location Description | Arrest |
|---|-------|-------------|------------------------------|-------------------------|------|--------------|---------------------------|----------------------|--------|
| 0 | 25021 | JD204167 | 04/07/2020 05:09:00 PM | 002XX S STATE ST | 0110 | HOMICIDE | FIRST DEGREE MURDER | CTA "L" PLATFORM | True |
| 1 | 23431 | JA340827 | 07/09/2017 02:36:00 PM | 000XX W 95TH ST | 0110 | HOMICIDE | FIRST DEGREE MURDER | CTA "L" PLATFORM | True |
| 2 | 24644 | JC347217 | 07/13/2019 03:10:00 PM | 001XX W CERMAK RD | 0110 | HOMICIDE | FIRST DEGREE MURDER | CTA "L" PLATFORM | True |



Data Cleaning

- Renamed columns to match style
- Dropped duplicate dates
- Changed all dates to the same format for each dataset
- Removed 2020 data for each dataset
- Deleted columns from crime data that were not used for analysis
- Merged datasets on date
- Replaced mislabeled rows
- Dropped crime data that was not CTA 'L' related



Data Analysis Methods

- Based on research questions, manipulated dataframe
- Choose appropriate plots for each question

Results: Rider totals and Crime Totals

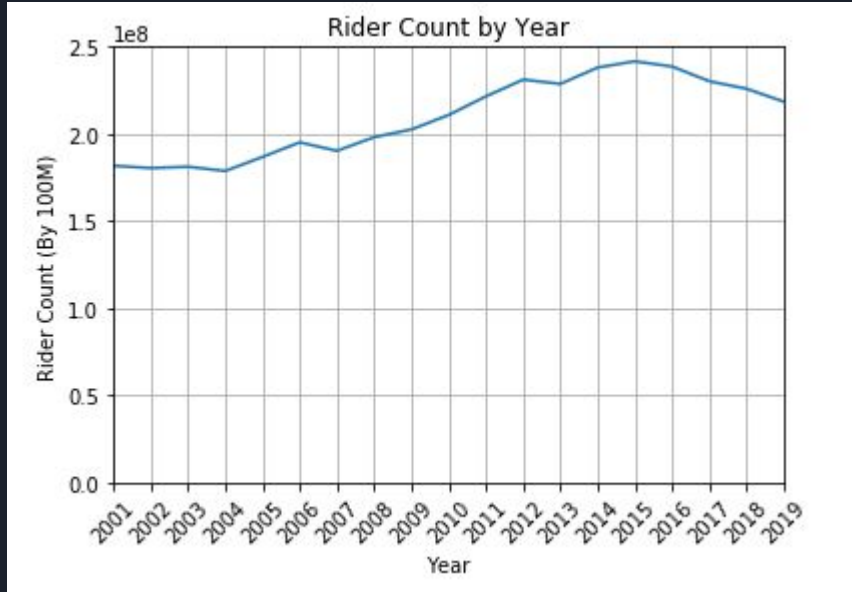


Figure 1: Total CTA riders per year

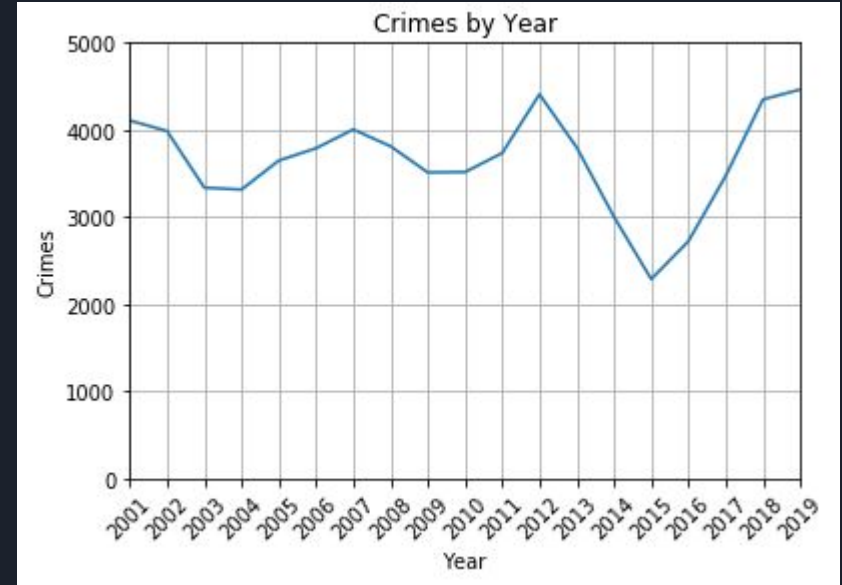


Figure 2: Yearly sum of crimes 2001-2019. Though the *number* of crimes peaks in 2012, the *rate* of crime, (Figure 2) is lower than 2007 and 2001.

Results: Crime Rate

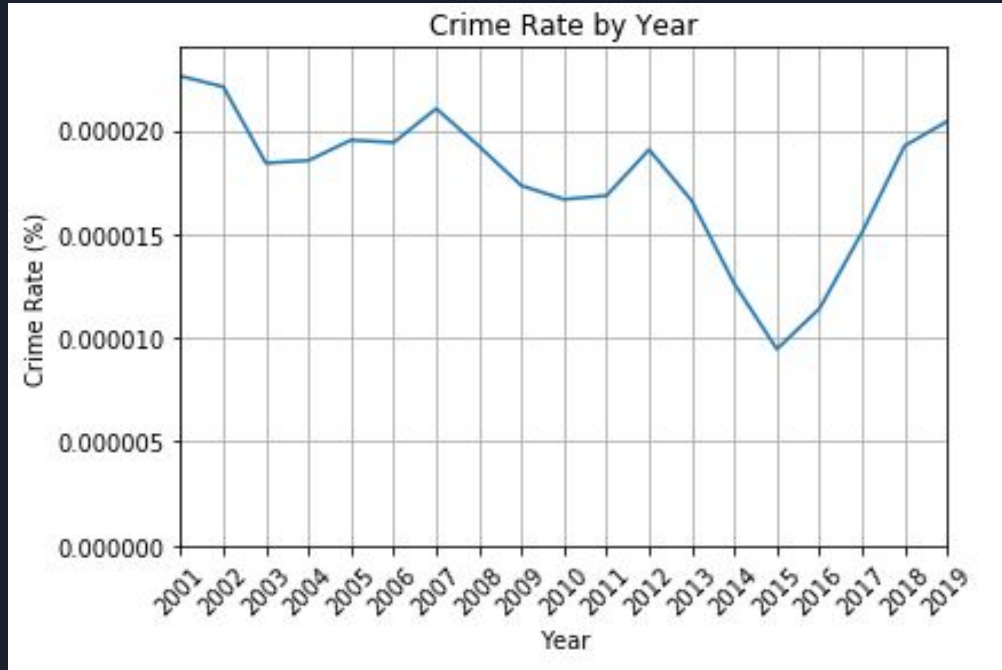


Figure 3: Crime rate: total crime in a divided by ridership. Highest crime rates were recorded in 2001, peaking again in 2007, and then a sharp decline 2012-2015.

Results: Crime Types

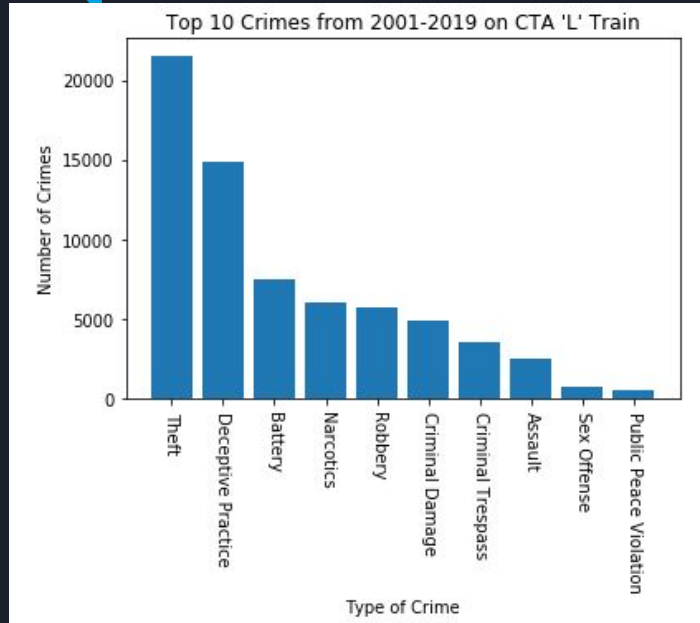


Figure 4: Top crimes by crime description. Top 5: Theft, Deceptive Practice, Battery, Narcotics

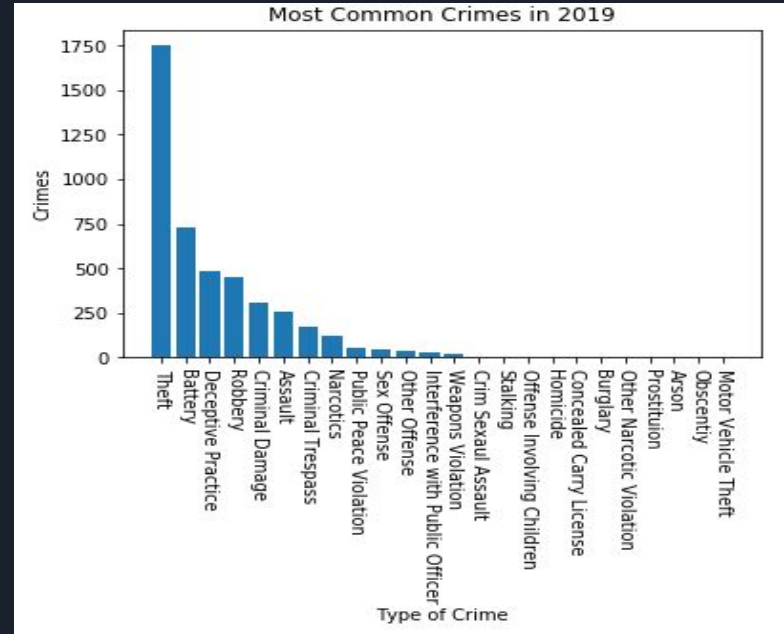


Figure 5 (Left): Most common crimes in 2019 by description

Most common crimes in 2019

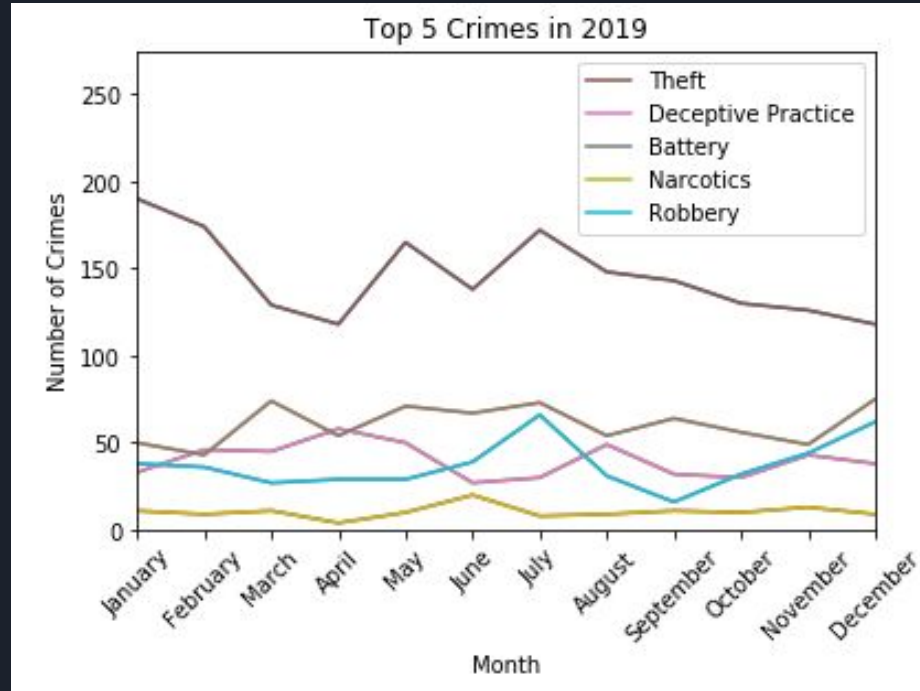


Figure 6 (Above): Top 5 crimes in 2019 throughout the year

Results: Crime on Special Days

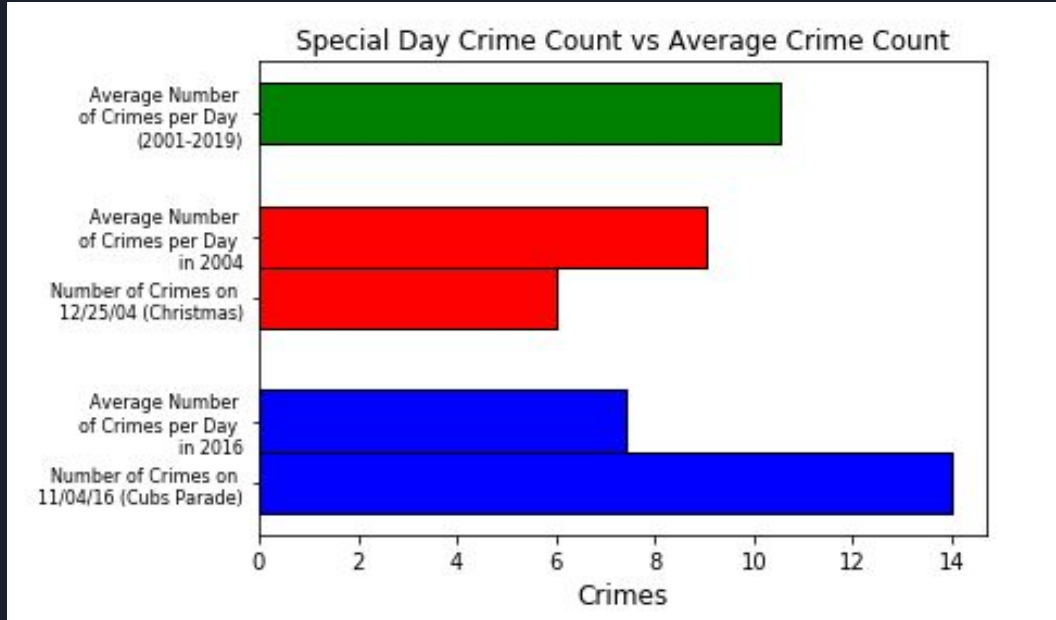


Figure 7: 12/25/2004: Christmas Day 11/04/2016: Cubs World Series Parade

Results: Scatterplot

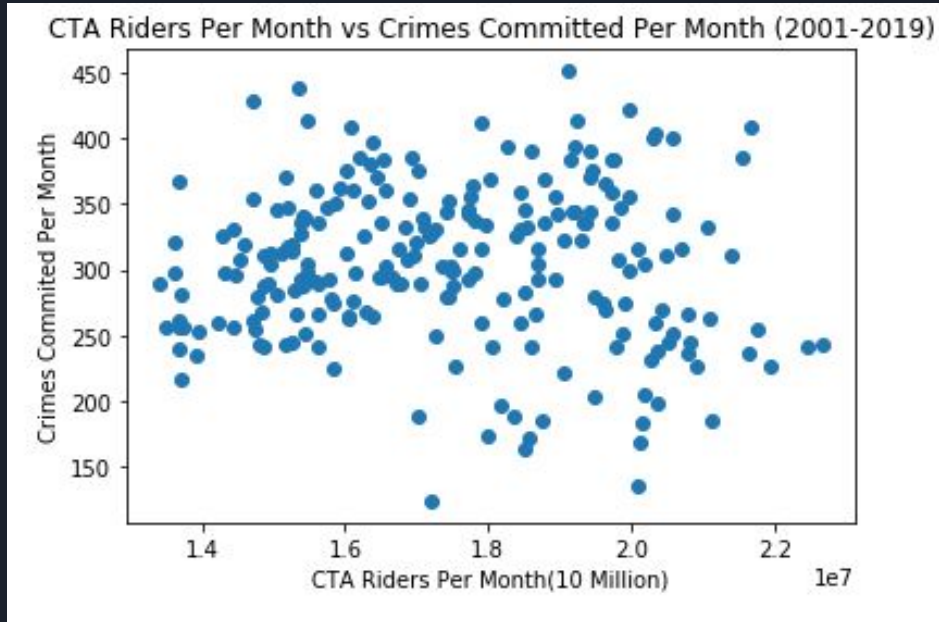


Figure X: Scatterplot displaying CTA Ridership vs Crimes Committed per Month. The correlation coefficient between CTA 'L' Ridership and CTA 'L' Crime is -0.04.



Discussion and Limitations

- Breaking down data further into route by route could provide more insight into correlation
- Many possible sources for rise and decline of crime and ridership numbers
- Difficulty in isolating date vs crime in the short term (ex. How does ridership change in the days after a publicized violent crime?)



Conclusion

- There is not a strong proven correlation between crime and CTA ridership over the period of 2001-2019, but both crime rates and ridership has changed over time
- Rising amounts of crime might be attributed to increasing ridership (more people = more crime) but the rate of crime does not consistently increase
- The rate of crime has a sharp decrease 2012-2015 and a rise 2015-2019



Further Steps

- How do these results compare to Chicago crime as a whole?
- What caused the decline in crime following 2012 and spike in 2015?
- How much do the cameras deter crime?
 - Data questions: possibility that police were incentivised *not* to list CTA as location
- What caused decrease number of CTA riders in 2014?
- Are some routes disproportionately affected by crime?

Questions?

