REPORT ON THE IMPACT OF COVID-19 ON TRAFFIC CRASHES IN CHICAGO FROM FEBRUARY TO APRIL 2020

BY

KEVIN LUU & YUE YU

NOVEMBER 17, 2020

MSDS 430: PYTHON FOR DATA SCIENCE

INTRODUCTION

Chicago is one of the most famous cities globally, with nearly 3 million people living here. As COVID-19 spreads around the world in early 2020, Chicago is not immune. We live in the

Chicago area, and any changes here will affect our daily lives. As one of the cities with the most comprehensive traffic, we intended to start with traffic accidents by comparing crash data to the same time as last year and find the impacts that COVID-19 had on Chicago.

We have taken raw data from the Chicago Data Portal. This dataset contains every recorded traffic crash on city streets within the City of Chicago limits and under the Chicago Police Department (CPD). Since September 2017, there have been no specific data entry problems, so we believe the 2019 and 2020 figures are relatively reliable. We only took data from February to April 2020 and February to April 2019 for comparison reasons. We chose this period because it is uniquely representative. The outbreak was evident in the United States in February 2020. The Governor of Illinois issued the Stay at Home order in March 2020, and the situation deteriorated in April 2020.

We hypothesize that people who do still drive will be more reckless thus causing more accidents, and the total number of traffic accidents will increase proportionally since we believe that in total there will be fewer drivers so we took that into consideration by later comparing them by percentages rather than counts. Finding unexpected effects is also the purpose of this analysis. Hopefully, this report's findings will help people who live in Chicago or work in Chicago's transportation industry.

DATA PREPARATION

We chose to compare the same time period from both years. February to April in 2019 had 27,818 crashes recorded and in 2020 had 20,118. After cleaning up the original data of 2019, we were left with 27,743 accurate records. For 2020 data there were 20,060 left. We removed many of the null cells and created consistency and the only off-case was in the column for hit and run where it was left blank when it was not a hit and run.

In the data, there was a total of 49 columns, not all of them were helpful for our analysis, so we chose to keep 17 of them which included Crash Record ID, Crash Date, Posted Speed Limit, Traffic Device Condition, Weather Condition, Lighting Condition, First Crash Type,

Trafficway Type, Roadway Surface Condition, Road Defect, Hit and Run, Damage, Date Police Notified, Primary Contributory Cause, Second Contributory Cause, Location, and Injuries Total.

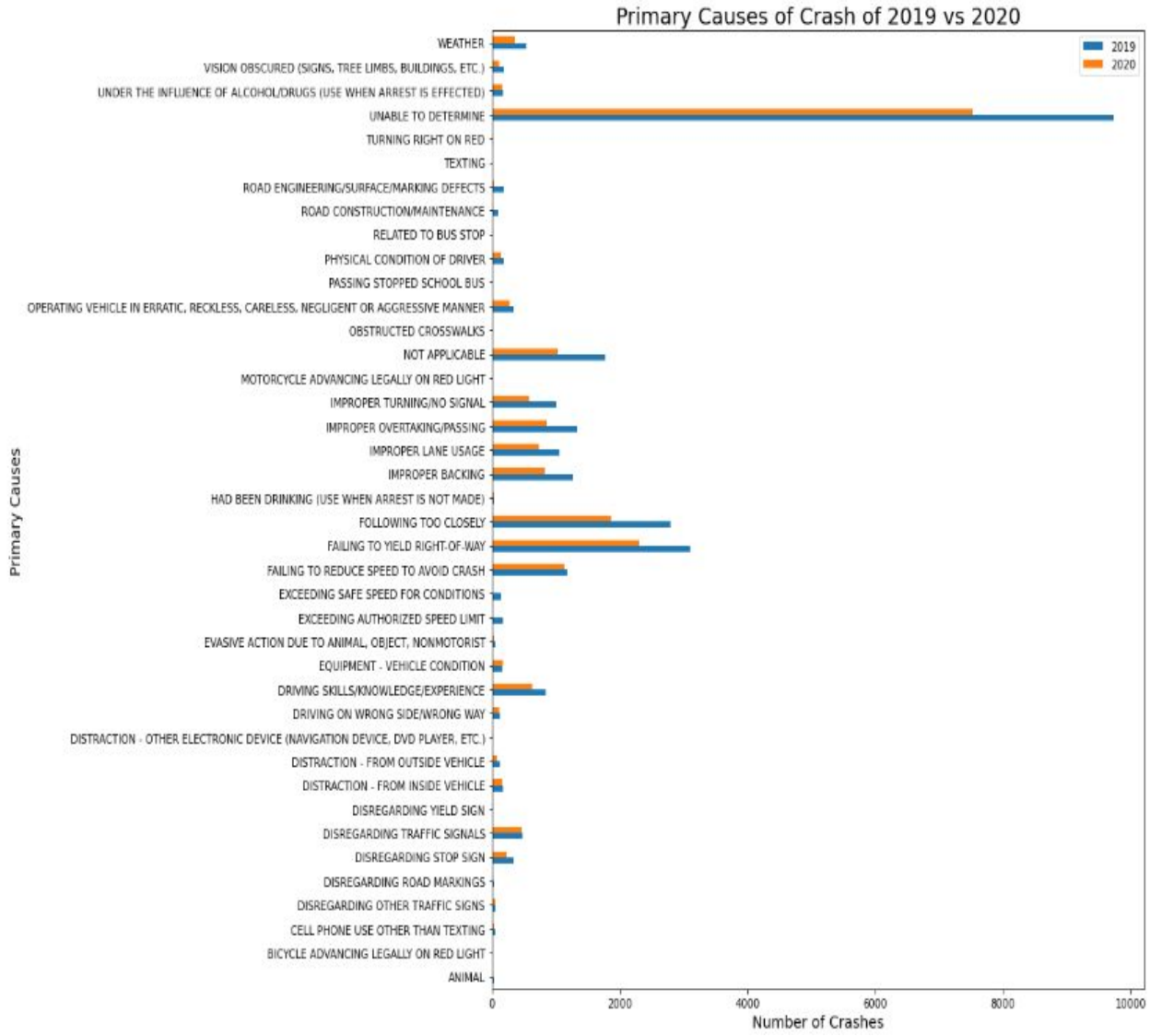
The next step is to clean up the invalid data in the retained items. A total of 27,818 records were recorded from February to April 2019, among which 19,833 had no accurate records of HIT AND RUN, and 74 had no correct records of Injuries Total. Because HIT AND RUN is not a necessary recording condition, but only a reference, and deleting nearly 20,000 pieces of data will have a drastic impact on analysis, we choose to keep it. We have cleaned up the records without Injuries Total. After cleaning up the original data of 2019, we were left with 27,743 accurate records. From February to April 2020, there were 20,118 crash records. After cleaning, there were 20,060 left.

ANALYSIS

Comparing Primary Causes of Crash of 2019 vs. 2020

We compared the data through multiple charts. We had to make some assumptions about the causes of crashes and or reasonings behind them. For example, the total numbers of reports were reduced in 2020 due to police numbers being reduced during a stay at home mandate.

Our first data visualization proved to show that 2019 clearly had more accidents, but we realized it would be incorrect to compare them by count since there were fewer drivers and crashes in total to be compared to last year so we switched to percentages. we took the count of the variable divided by the total count of crashes for that year, to the percentage for that year, for example, we would have data that would tell us 15% of crashes in 2020 were due to “failing to yield right of way” there would do the same for 2019. We did this so we can determine if a category or reason went up when compared to the previous year. This led to interesting findings.

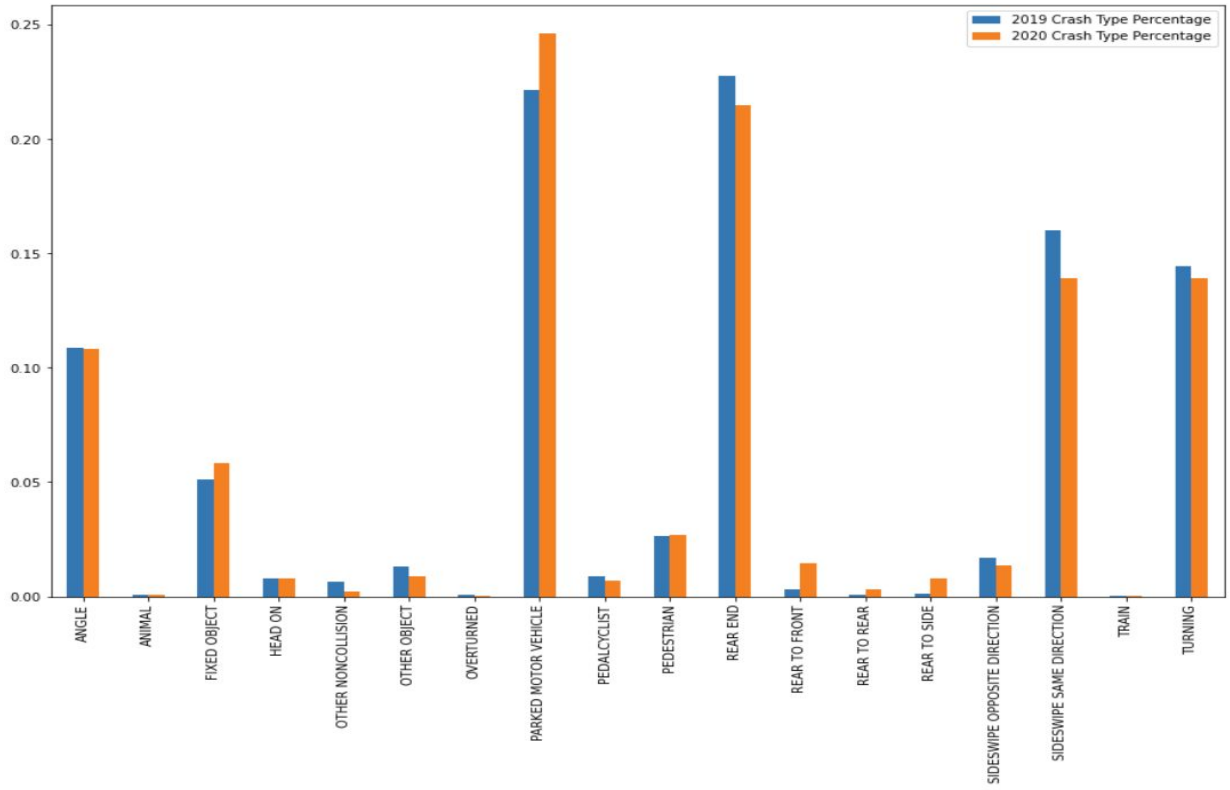


Converting the chart to comparing the percentages

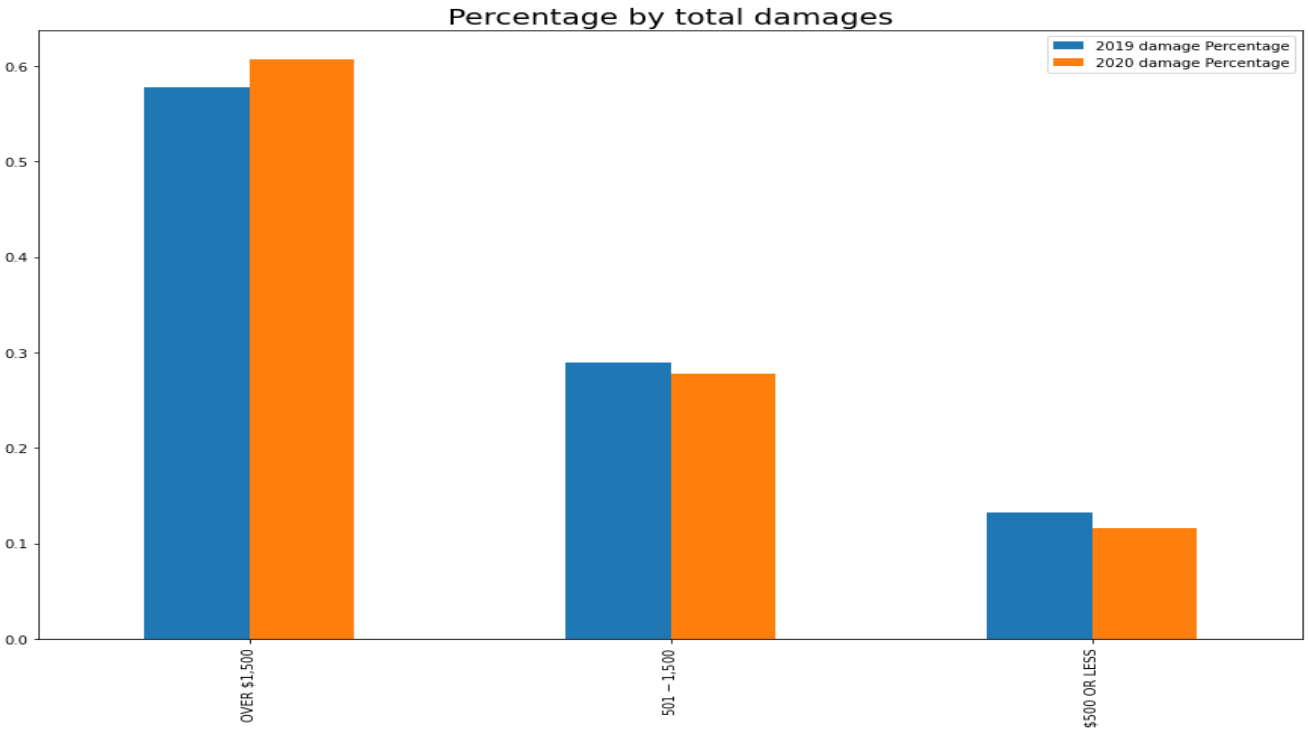
After looking at the data, we could not indicate driving behavior changes when comparing 2019 and 2020. but with percentages, we were able to see where driving behavior had changed. We found reasons like "failed to yield right of way" have gone up, and if we solely depended on the count instead of percentage, we would not have gotten the full picture and made a wrong assumption.

Crash Type Comparison by percentage for that year.

Since the data had many unknown reasons for crashes, we decided to then analyze the data from the crash type perspective using the same percentages method.



We were not expecting to see "Parked motor vehicles" go up; we were hoping to see categories like "Head On, rear end, side-swipe to go up as those seem to be more dangerous and clearly support our hypothesis. We were on the right track, but we can't make a clear call based solely on the type of crashes; maybe we should gear it towards the total damages cost to see if drivers drove faster than usual and see if locations change. Looking into damage cost changes could help us determine if drivers were driving more recklessly or overall driving fast because the faster you drive the crashes will be more severe.

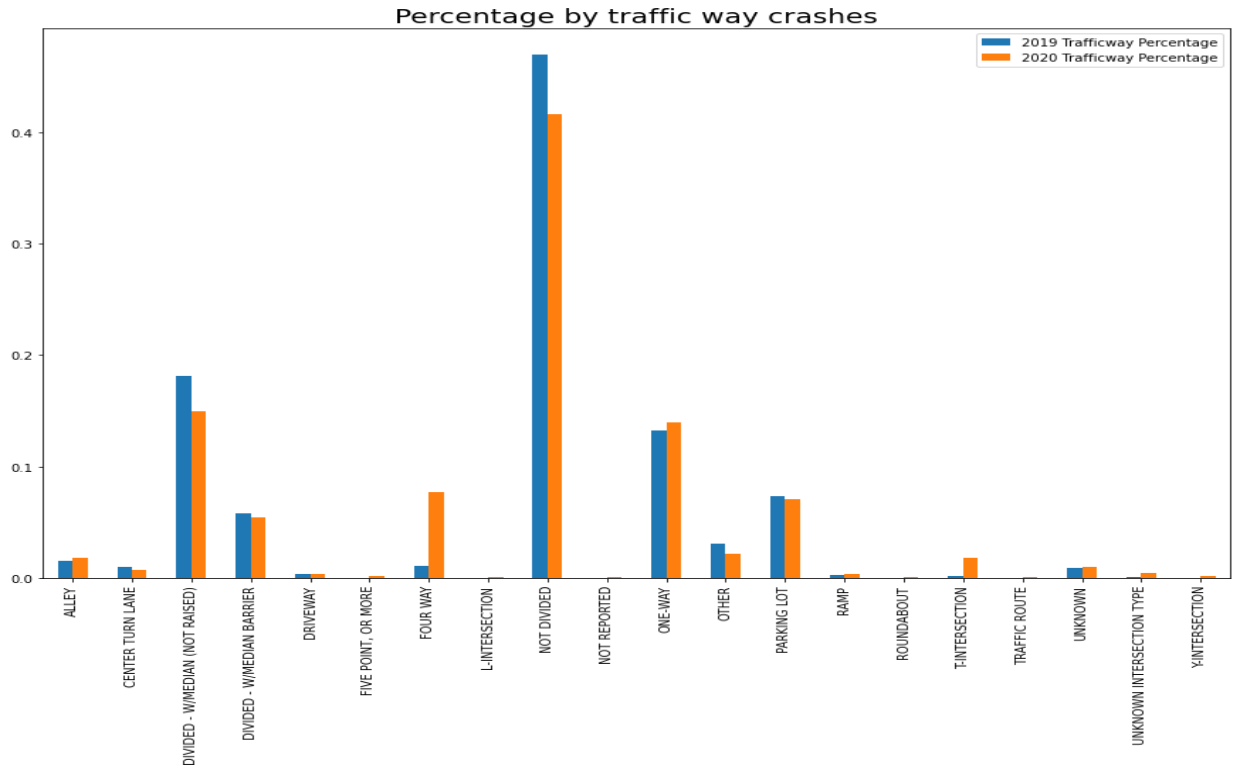


Here we can see that there has been a 3% increase of significant damages with "Over $1,500" This is as good as we can get in terms of severity of 2020 crashes because the faster or more reckless they drive the more expensive it becomes.

We also can assume that these increased damages are happening to fixed objects. Things like "Park vehicles" because as from the other chart we can see that they had a 2% increase as well as a few other “Non-moving objects”. This makes sense because even if the vehicle were speeding or driving recklessly, hitting another car would be less likely due to fewer drivers on the road. So non-moving object-related crashes have increased.

Traffic Way Type

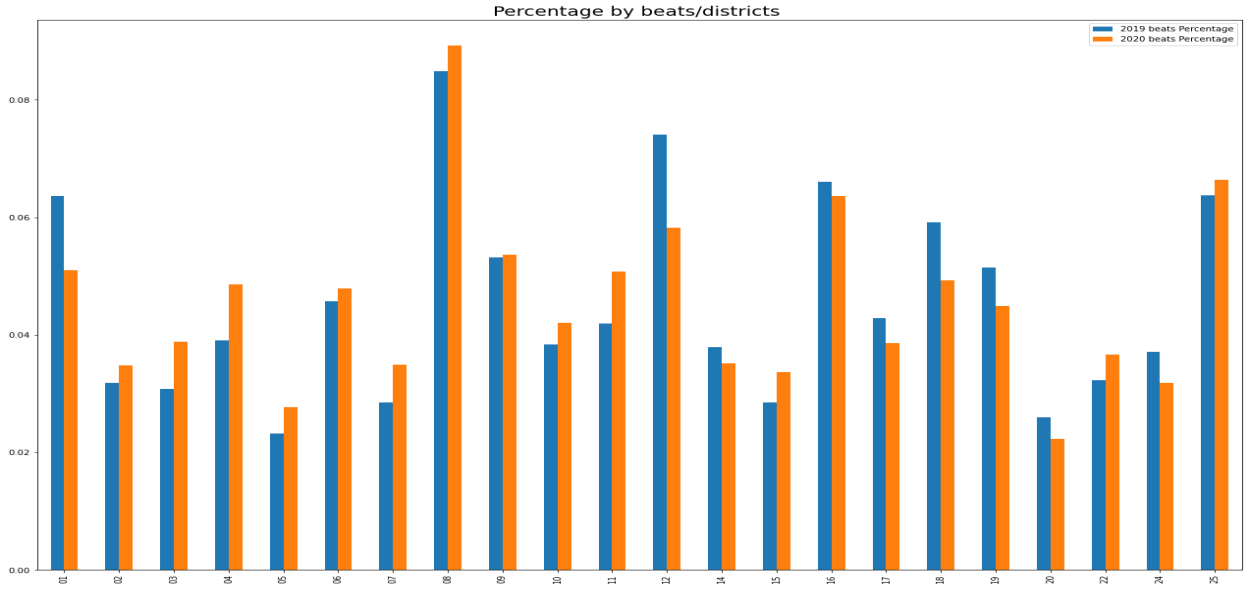
Traffic Way is where the crash happened, like intersections, one-ways highways, etc. taking into consideration that fixed object crashes increased. We were expecting crashes to more likely occur in areas where more things could get hit, like narrow streets,one-ways, and non-divided roads. This was true for the most part so now we can see that this supports our discovery.



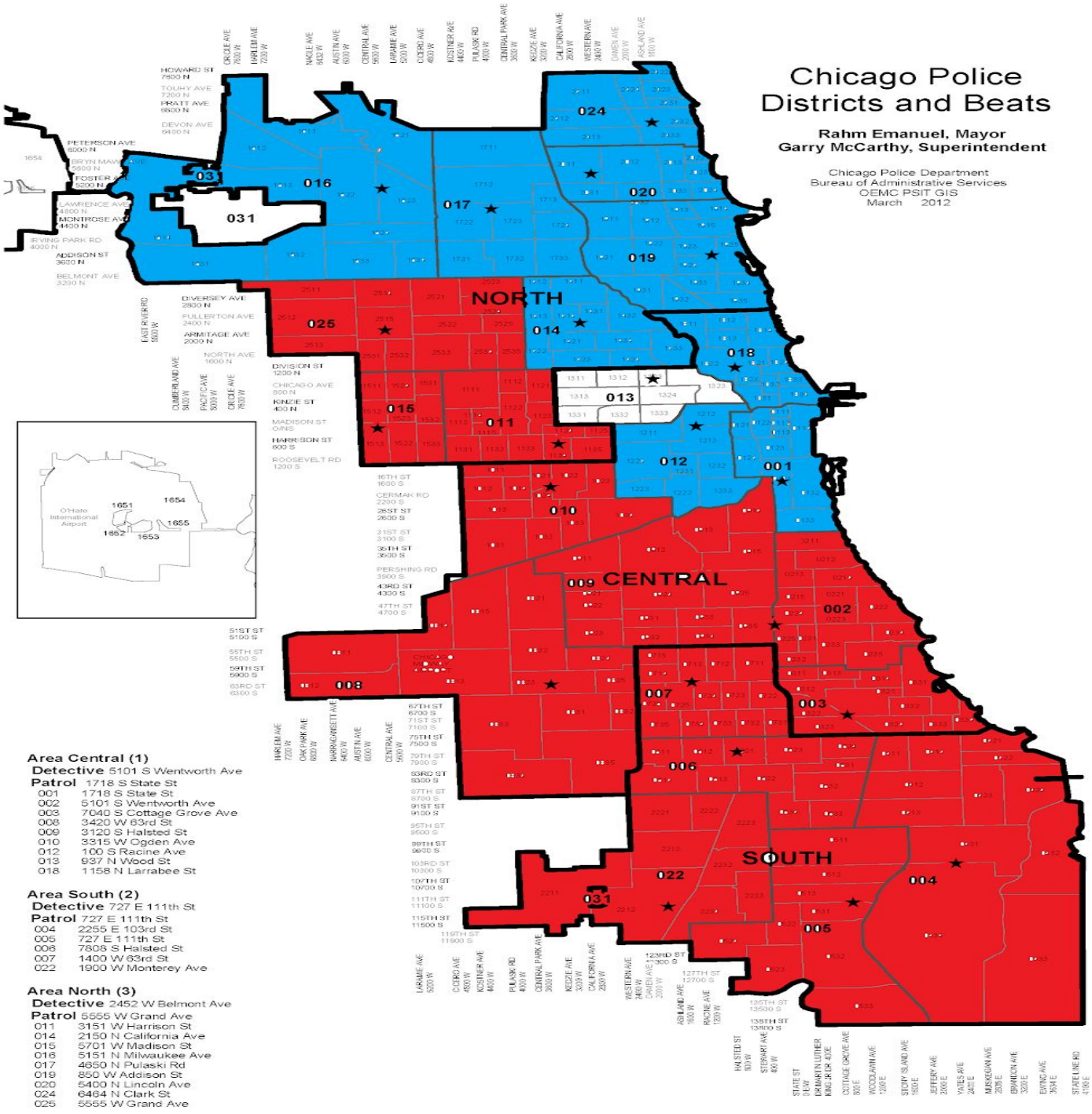
We were beginning to understand that crashes did increase when compared to 2019 but not exactly the way we were expecting. instead of it being a car on car crashes it ended becoming crashes to fixed objects which would explain the crash reasons like “Failing to slow down to prevent crash and traffic ways where there are likely more objects to hit.

Understanding Percentage by beats/districts

Next, we wanted to analyze the Chicago area to see if there was any change in each district's number of accidents. We were not sure what this could uncover since it would be hard to determine if certain areas had more fixed objects than another but it could give us important information on where these crashes would most likely occur.



We can see that there has been an increase of 13 beats and a decrease of 9 but what is more interesting, is that all increases in crashes occurred on the south side. Simultaneously, the Northside has decreased, while the central Chicago remains somewhere in between. We think it's safe to assume that when crashes did occur, it was mainly happening in southside > central > northside. We show the trend of regional accidents by staining the area. Blue indicates fewer crashes, and red indicates an increase in the number of crashes. As shown in the figure below:


Chicago Police Districts and Beats
Rahm Emanuel, Mayor
Garry McCarthy, Superintendent
Chicago Police Department
Bureau of Administrative Services
OEMC PSIT GIS
March 2012
NORTH
CENTRAL
SOUTH
001
002
003
004
005
006
007
008
009
010
011
012
013
014
015
016
017
018
019
020
022
024
025
031
Area Central (1)
Detective 5101 S Wentworth Ave
Patrol 1718 S State St
001 1718 S State St
002 5101 S Wentworth Ave
003 7040 S Cottage Grove Ave
008 3420 W 63rd St
009 3120 S Halsted St
010 3315 W Ogden Ave
012 100 S Racine Ave
013 937 N Wood St
018 1158 N Larrabee St
Area South (2)
Detective 727 E 111th St
Patrol 727 E 111th St
004 2255 E 103rd St
005 727 E 111th St
006 7808 S Halsted St
007 1400 W 63rd St
022 1900 W Monterey Ave
Area North (3)
Detective 2452 W Belmont Ave
Patrol 5555 W Grand Ave
011 3151 W Harrison St
014 2150 N California Ave
015 5701 W Madison St
016 5151 N Milwaukee Ave
017 4650 N Pulaski Rd
019 850 W Addison St
020 5400 N Lincoln Ave
024 6464 N Clark St
025 5555 W Grand Ave

CONCLUSION

Before starting the analysis, we predicted some results, but not all of them. Compared with the same period in 2019, the total number of crashes in 2020 is significantly reduced. But we suspected that increased speeding/reckless driving would occur thus resulting in more injuries and deaths, etc. but we learned that car-on-car crashes have gone down, but crashes that hit property or fixed objects have gone up. The crash location has increased in one ways and 4-way intersection where drivers either hit parked cars or speed through the intersection. The further south we go the more likely there are crashes related to these scenarios. and the damage cost has gone up so we can make the assumption that drivers were driving more recklessly during the stay-at-home mandate.

Due to the Stay at Home order's introduction, the proportion of vehicles hitting objects has increased compared to other vehicles. And the proportion of losses caused by car accidents also increased compared to the same period last year. Serious accidents are more likely to occur during Stay at Home order. We attribute this to drivers being careless and committing serious offenses when there are few people around.

The percentage of changes between 2020 and 2019 were slight. So one thing that we could have done to increase data quality was to pull multiple years and taken the average and then compared it to 2020 to hopefully get a more accurate percentage change. Most of our issues occurred during part 2 when we discovered our hypothesis was incorrect since car on car crashes did not occur and took a while to determine an actually change till we chart/graphed everything.

In summary, the positive impact of home orders on car crashes is clear. In addition to the reduction in total volume, vehicle-to-vehicle collisions have also been reduced. Drivers' personal and objective circumstances cause more crashes. There are apparent differences between regions, which should be analyzed in detail in regional enforcement of decrees. Finally, beware of drivers who engage in dangerous driving during this period. The analysis result shows that the proportion of these people has increased over the past year.