

Traffic in Chicago

Team 12: Nick Matthews, Jian Lee, Meredith Tucker, Mike Zaino

Chicago is one of the most populated cities in all of the United States. Due to this high population, the roadways endure high traffic patterns from the daily travel of people. No matter how careful people are, there are always going to be accidents on the road. This could be caused by anything from human-error to road conditions to weather-related reasons.

The crash dataset provided by the city of Chicago obtained contains 189,724 entries with 49 columns. This data set included categories such as the date of the accident, weather conditions, longitude and latitude of the location of the accident, the speed limit of the road, and many more to describe the accident. This data was acquired from cityofchicago.com. After acquiring the data, the first set was cleaning the data to fit the needs of our analysis. We decided to limit the total number of entries to allow data analysis to be smoother due to a large number of entries in the complete dataset. We took the date category and first had to separate the column into month, day, and year. We then filter the data to only be the year 2016. This was chosen because 2016 had the most entries and was the latest year in the dataset. The data with the years 2013 and 2014 had fewer data entries and 2017 did not have entries for all months.

The four questions we will analyze are the specific date with the most accidents and whether it is related to any activity, the impact of weather on the accidents occurring in Chicago, the most common primary and secondary factor in crashes, the number of injuries in each primary factor, the top 3 primary factors occurred in what time and visualization of the locations of each crash change from month to month.

Days with Most Accidents in Chicago

The first goal was to find the days with the most amount of accidents. This was achieved by grouping the dataset by date. Since we changed the date into the month, day, and year, we had to group by all three of those variables. From there, we totaled each entry with a given date.

Date	Total Accidents.
12-19-2016	260
12-20-2016	258
11-23-2016	235
10-26-2016	234
12-16-2016	220
12-14-2016	216

From these, we can see a pattern. All of these dates with the most accidents were during Winter months.

The next thing to find was what factors were causing many accidents during these days. Instead of looking at the dataset, we decided to research what was going on during these days around the accident. We looked at sports games, the weather, and anything we believed that could affect traffic to cause more accidents.

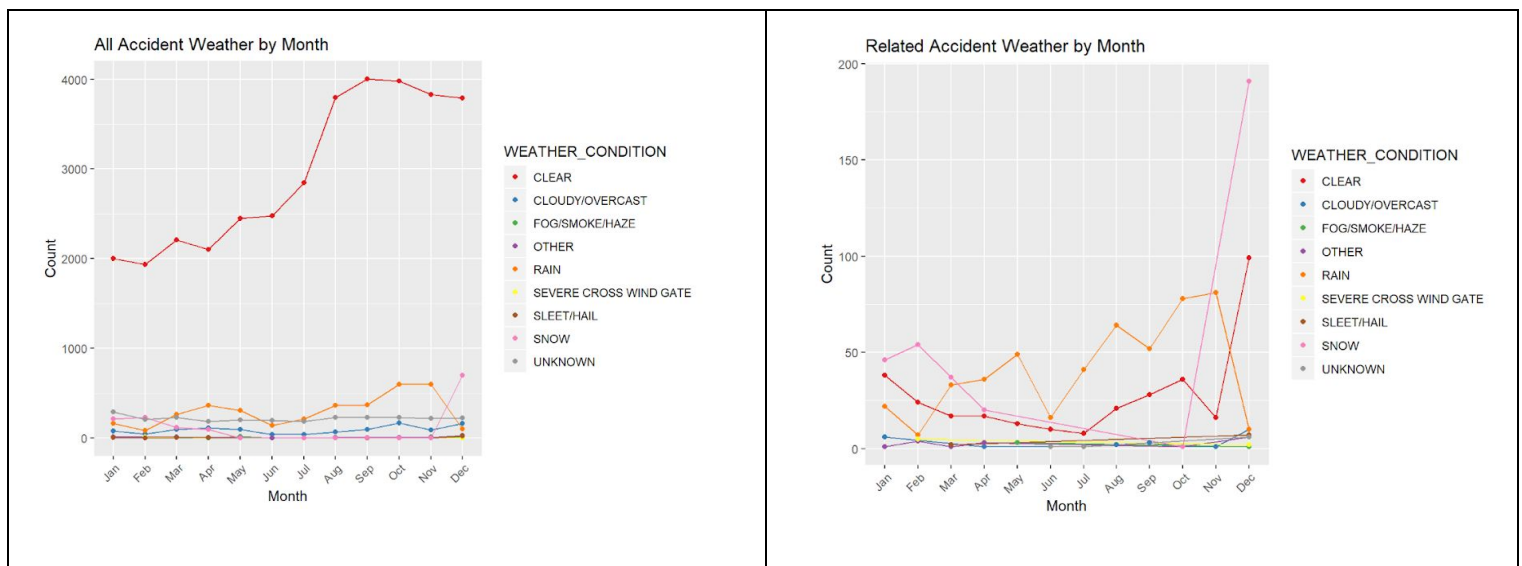
Date	Sporting Event	Weather Event	Other Events
12/19/16	Pistons vs Bulls	High of -6°F	12/18/16, Wind Chill Advisory, First week out for Chicago Public Schools
12/20/16	Senators vs Blackhawks	High of 0°F	12/18/16, Wind Chill Advisory, First week out for Chicago Public Schools
11/23/16	Predators vs Blackhawks	Rain All Day	11/24/16, Thanksgiving Day Parade
10/26/16	Spurs vs Bulls	Rain All Day	NA

12/16/16	Islanders vs Blackhawks	Thick Fog	NA
12/14/16	NA	2 inches of Snow	2 inches of Snow day Before

Searching for specific events on a given day was made easier because we were trying to find events specifically in the Chicago area. Looking at the table above, we can see that on all of these days, there was a sporting event taking place. Because the Bulls and the Blackhawks play in the same stadium, we can pinpoint their locations to the United Center. Next, is the type of weather that was going on during those days. We wanted to get as specific as possible. During the two most populous days, the temperature was at or below 0°F for the majority of the day with a wind chill advisory. On top of that, public schools in Chicago were on their first week of Winter break putting more people on the road and traveling. Next, on November 23rd was the day before the Thanksgiving Day Parade. We believe that there could be more accidents on this day because Chicago has to close down certain roads for the parade to take place. As a result of these closures, more cars will be on fewer roads causing more traffic which could lead to a higher chance of an accident occurring. Finally, four of the top six days listed are all within a week from each other. The cause behind this observation is most likely the weather. However, another factor could be the Holiday season. Around December, schools and businesses are closing for Seasonal Holiday breaks, leading to an increase in travel. The most popular mode of transportation in the United States is driving. Chicago is no exception. With more cars on the road, there is going to be a higher chance of risk of accidents.

Impact of Weather on the Accidents Occurring in Chicago

When analyzing this data set in regards to the effect of weather, it was necessary to distinguish the accidents that listed weather-related factors as a primary or secondary cause. The data was then reduced to the accidents that listed the causes of the accident as either weather or exceeding safe speed for conditions. Accidents that were contributed to weather-related causes accounted for 2.79% of all accidents that occurred. While overall, weather-related factors may seem to have a relatively insignificant impact on the cause of accidents, the weather and roadway surface conditions provide greater insight. The cause of 41.2% of accidents in windy conditions was listed to be weather-related. For accidents in snowy conditions, 25.6% of accidents listed weather as a cause. Accidents occurring in conditions of hail and sleet attributed 20.0% of accidents to weather. The condition of the roadway surface also showed a relation to the cause of accidents. The accidents with icy road conditions showed that 41.5% of these accidents were due to weather-related factors. The number of accidents per weather condition was plotted by month to show the influence of the season.

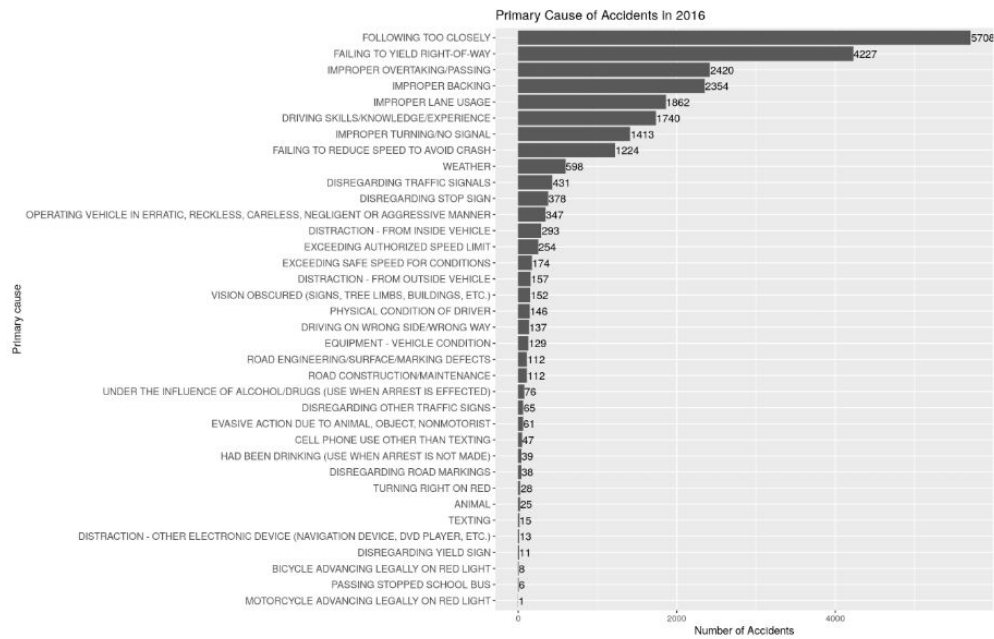


The most notable difference between the graph concerning all accidents and accidents due to weather-related causes were the accidents that occurred on clear days. The amount of accidents that occurred in clear weather is reduced dramatically in the data relating to weather

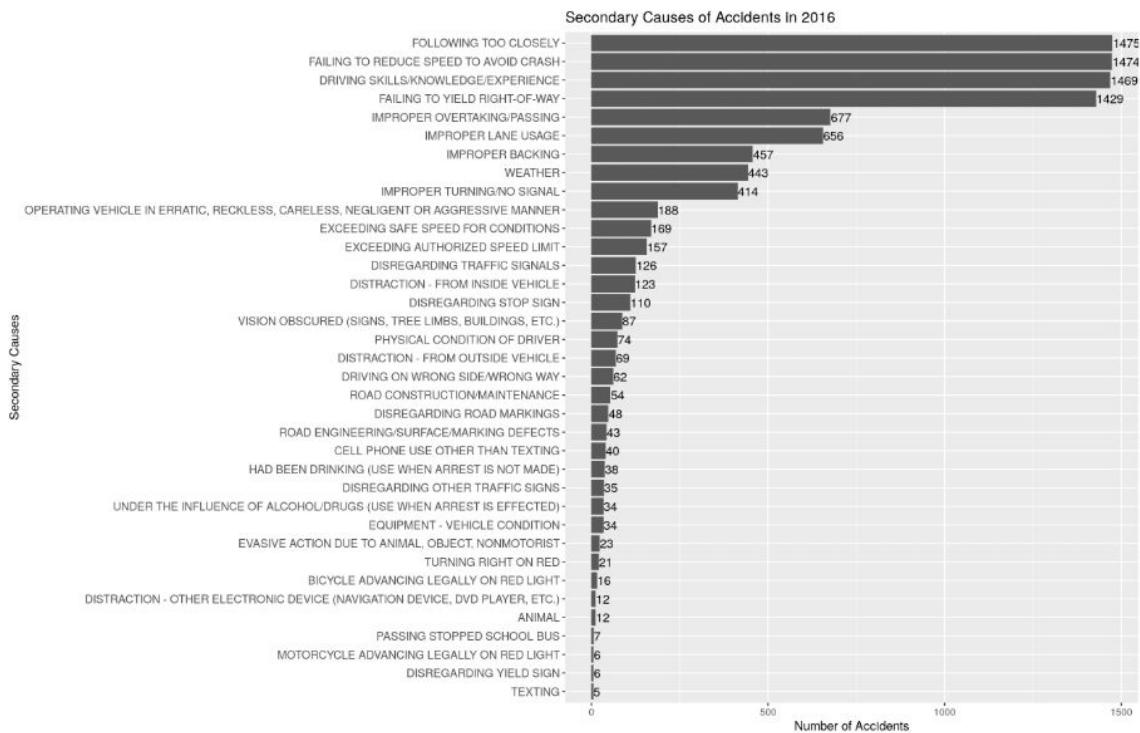
conditions. However, it still appears as one of the top three weather conditions for accidents due to weather conditions. The calculations as described above showed that 0.92% of accidents on clear days were contributed to weather-related conditions. This visual shows that although the percentage is small, the amount is not non-negligible with regard to the number of accidents occurring. The plots of the weather conditions that had large percentages above maintained a similar path in both graphs. The accidents that occurred in rain weather conditions consisted of 13.7% of the accidents to be caused by weather. The plots of the rain data appear to be similar but show several differences. The graph of all accidents shows a small peak in April, however, in the weather-related accidents graph this peak occurs in May.

Primary factor and Secondary Factor causing crashes

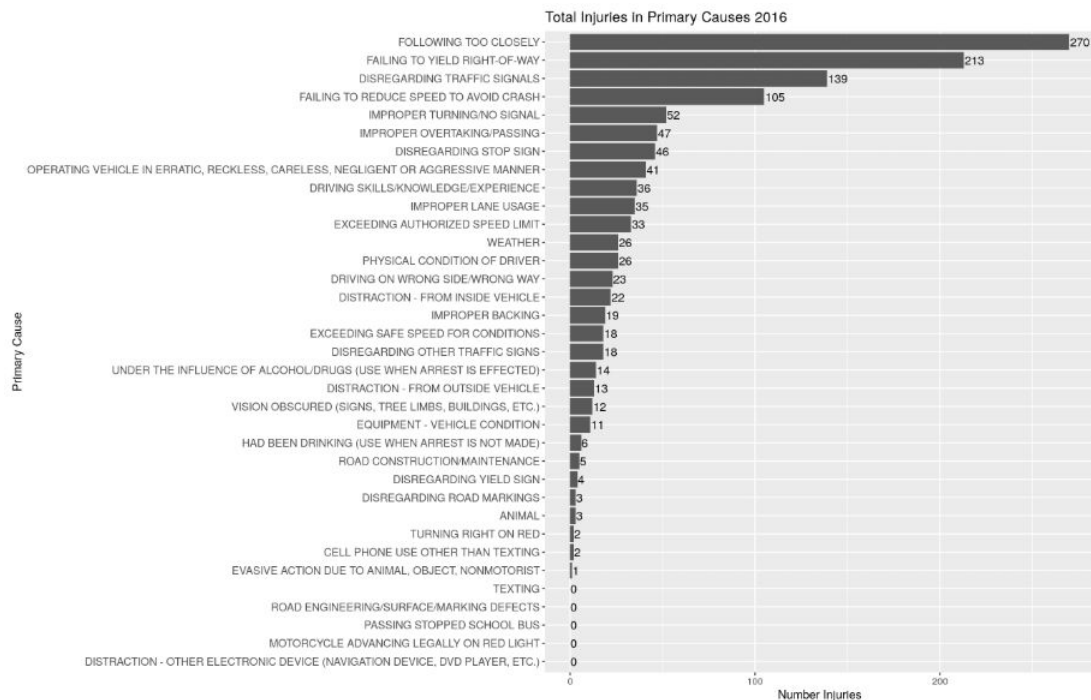
In the dataset, it provides the primary causes and secondary causes of the accidents in Chicago. It is important to analyze the primary causes and secondary causes contributing to the accidents. There are 36 factors in the primary and secondary causes columns in the dataset. As displayed by the barplot, the most common primary cause is following too closely with 5708 out of the total 24801 accidents. The percent of this cause in accidents is 23%. Therefore, following too closely is the primary cause of accidents in urban areas. The second most common cause resulting in 17.04% of accidents is a result of failing to yield the right of way. The third cause is improper overtaking or passing. There are 2420 accidents related to improper overtaking or passing.



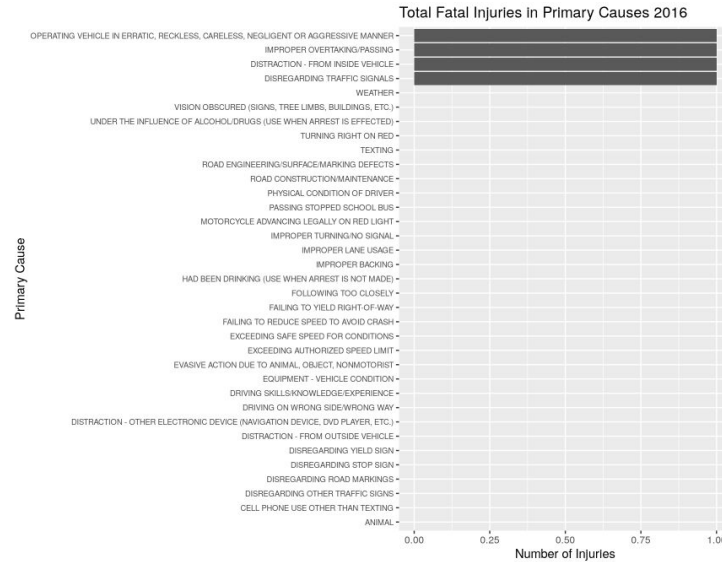
In terms of secondary causes, the first three most common are following too closely, failing to reduce speed to avoid a crash, and driving skills/knowledge experience. There are 1475, 1474, and 1469 accidents related to them, respectively.



When comparing the total injuries to the primary causes, the highest number of injuries occurred when the driver followed too closely. There were 270 injuries when the driver followed too closely. The second is failing to yield right of way. This caused 213 people to be injured. The third is disregarding traffic signals, resulting in 139 people injured. In an urban setting, like Chicago, following too closely while driving and failing to yield right of way result in higher chances to cause driver and passenger injuries. The percentage of these two causes in terms of the total number of injuries is 38.8%.

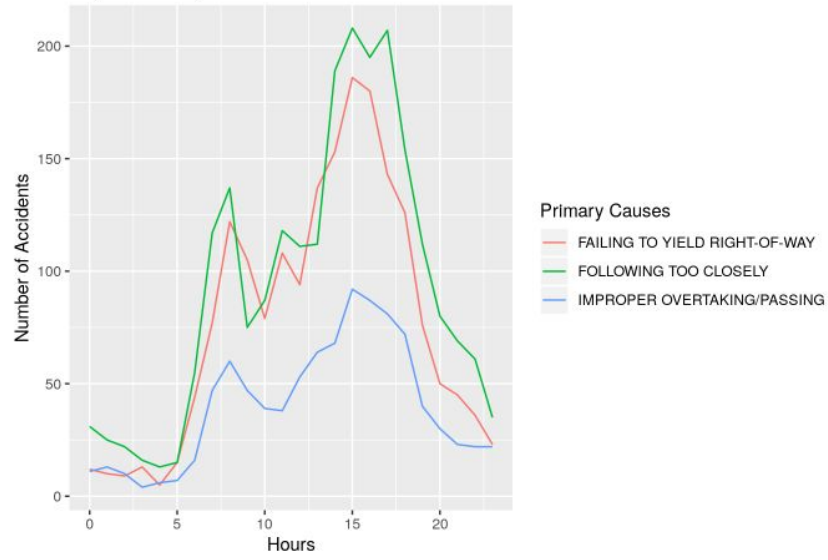


The total of fatal injuries in Chicago 2016 is only 4 injuries. The rate of fatal injuries compared to total injuries in Chicago is $4/1245 = 0.3\%$. The rate of fatal injuries in terms of total accidents is relatively small. The causes of these are driving recklessly, carelessly or aggressive manner, improper overtaking, a distraction from inside the vehicle, and disregarding the traffic signal. Therefore, driving in Chicago has very low chances to cause fatal injuries.



The line plot shown below displays the number of accidents of the top 3 primary causes occurring in a 24 hour period. When the hour is between 0-5, the number of accidents that are a result of the top 3 causes is the lowest in the day. There is no significant difference between these causes. When the hour is between 5-10, there is an increased number of accidents occurring. Following too closely causes the greatest increase. People tend to drive to work during these times when accidents occur. Between 15-17 hours is the highest number of accidents because during this time people are off their work and drive back to home. Therefore, the plot shows the accidents that occur when people go to work and off from work has an increasing pattern.

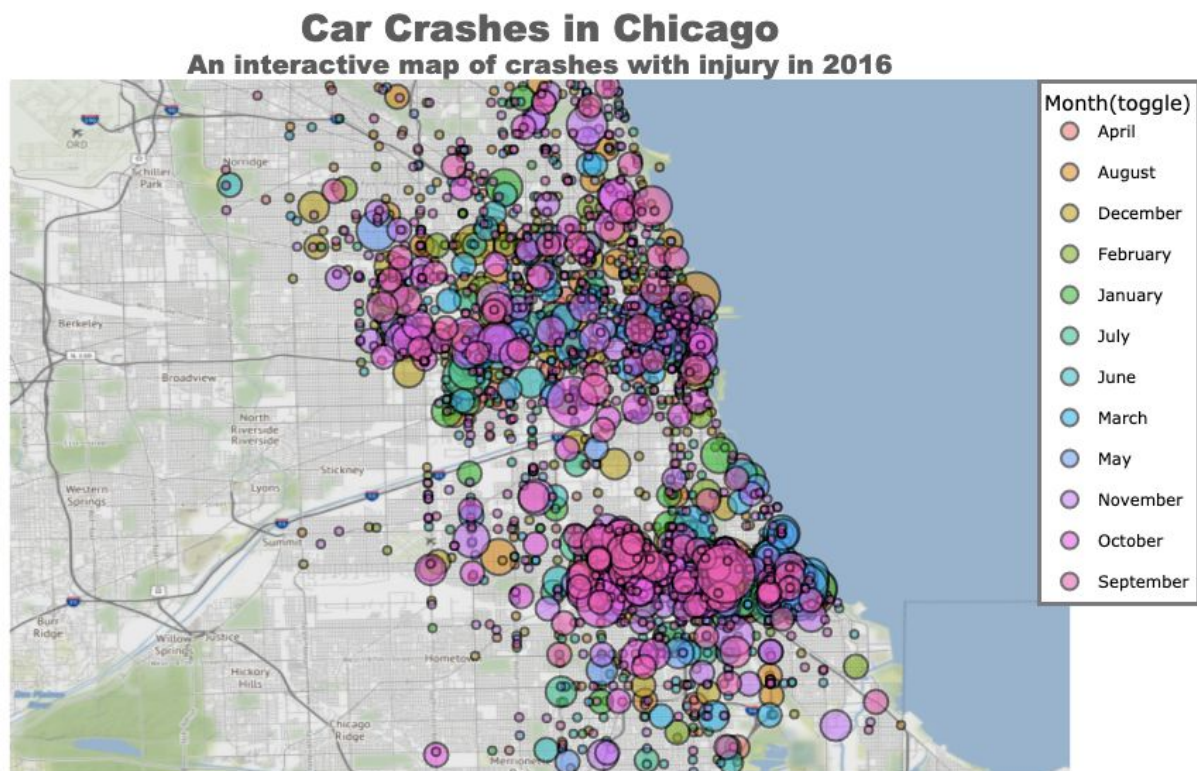
Top 3 Primary Causes of Accidents occurred in Hours



Visualizing the Locations of Each Crash Change, Month to Month

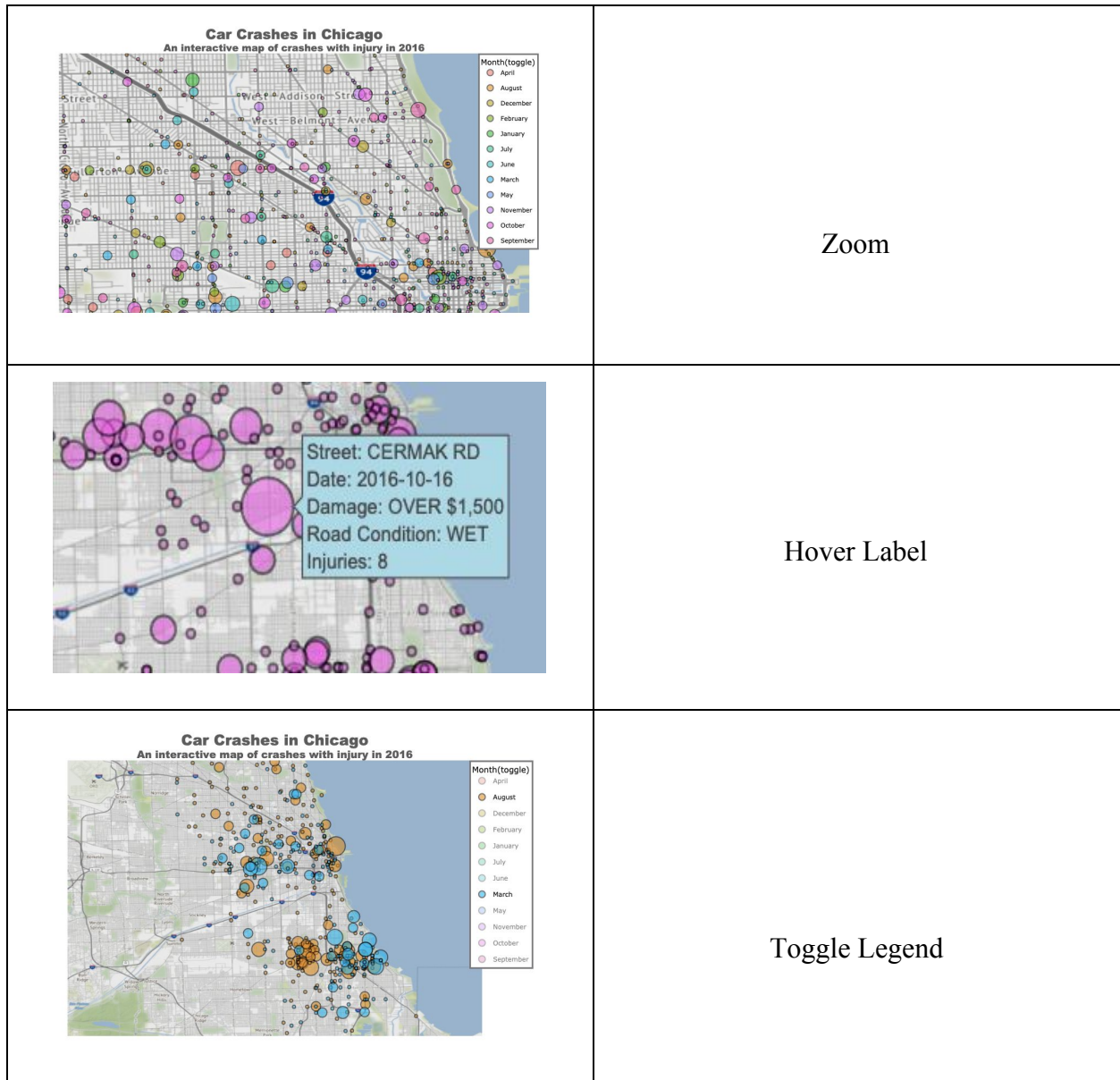
To display the location of each crash, we used the ggmap library which uses images of google maps to display a more detailed visual of Chicago. Then using the longitude and latitude of each crash, we place each crash with injury on the map. Next, to make the map more informational, we mapped the color of each point to the month of the year and the size of the point to the “severity” of the crash(number of injuries). Finally, to make each point distinguishable, we used theme elements to change the opacity, color, and border of each point on the infographic.

However, this was still too much data for someone to read effectively. To remedy this, we first created a subset of the dataset which only included crashes with 1 or more injuries. This made our visual easier to read and did not skew the data.



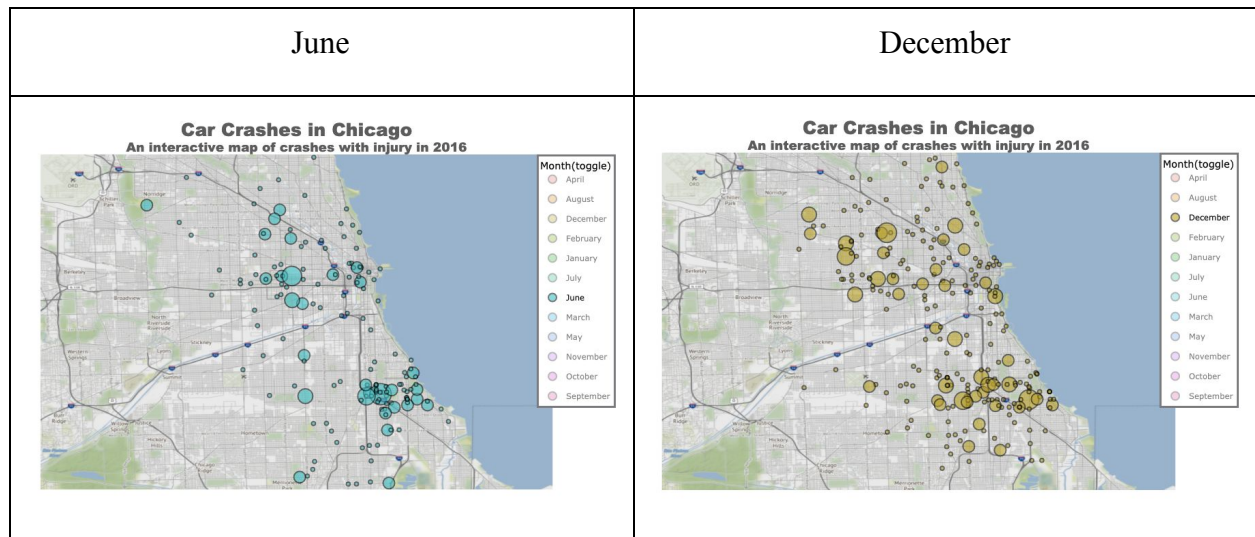
Though, once again, there's not much room for analysis here. The final step in making the map a readable tool for analysis was by adding 3 functions: Allowing zooming into the map,

displaying qualitative information as you hover over each point, and adding the ability to toggle the items within the legend to display any combination of crash months.



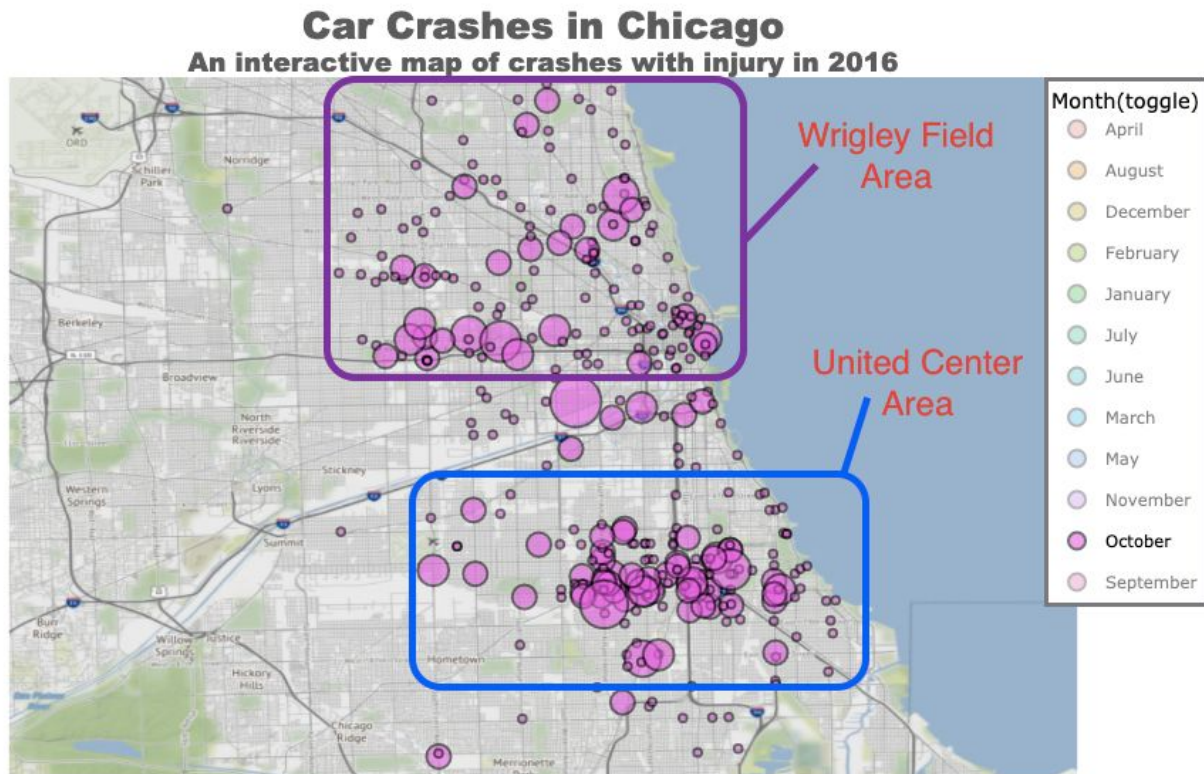
Once the tool was built, we were able to analyze it. We began by searching for common claims made about Chicago travel. Santorini Dave, who hosts a popular traveling information website, claimed that one of the best times to visit Chicago is in June and August because of the great weather and easier driving conditions. Well, we know that driving conditions have an

impact on crash numbers (intuitive), so let's put this claim to the test using the map by looking at June in comparison to a winter month when the roads are at their worst.



As you can see, when you toggle between June and December, there is a large difference between the number of crashes seen on the map.

There are many reasons why crashes might increase and change location through the months of 2016, though. For example, you can tell the exact month that games 3, 4, and 5 of the 2016 World Series were played in Chicago when clicking on October. There are lots of crashes near Wrigley Field, both low injury and high injury. October is also when the NBA season begins. So, as you can see near the United Center, which is one of the largest NBA stadiums in the world, there are many crashes right around the building itself.



Challenges

The group faced many different challenges individually. But, as a whole, one of the biggest roadblocks involved data cleaning. More specifically, scaling down the data. We proved challenging to bring the data down to a size that still gave us reliable analysis. For example, on the ggmap, plotting every single fender bender to occur in Chicago resulted in an enormous, unreadable infographic. It had to be sized down so that it was easy to read, but without skewing the data.

Another part of our process which proved difficult was finding the right way to display the final product. With many unique causes of car crashes and all sorts of road conditions, it was hard to put them on a single axis or in a single ggplot without lots of overlap. Though, utilizing good theme elements like angled axis labels and distinguishable colors made this less of an issue.

Conclusion

Our analysis of this data focused on four major parts, factors relating to the dates with the most accidents occurring, the effects of weather conditions, the contributing primary causes, and analyzing the data through mapping. We found that one of the factors relating to crashes is sporting events with increased numbers of incidence occurring on the dates of these events and that these accidents are located near the hosting establishment. It was also found that times with increased traffic patterns such as holiday breaks, or common commuting times result in more accidents occurring. Seasonal weather patterns were also found to play a role in traffic accidents through correlation of weather conditions and mapping techniques. Although many factors contribute to traffic accidents, we were able to identify some patterns relating to them. It is important to be aware of the factors that contribute to traffic accidents to create a more conscious general public to improve traffic safety.

Sources

<https://data.cityofchicago.org/Transportation/Traffic-Crashes-Crashes/85ca-t3if>

https://www.basketball-reference.com/teams/CHI/2017_games.html

https://www.hockey-reference.com/teams/CHI/2017_games.html

Dave, Santorini. "Best Time to Visit Chicago – The 2019 Guide." *Santorini Dave*, 6 Oct. 2019, santorinidave.com/best-time-to-visit-chicago.