Anna Poon and Kaya Lee
DATA 11900, Spring 2022

Chicago Car Crash Data 2016-2021

The City of Chicago's Traffic Crashes – Crashes dataset is a log of the roadway crashes that occur within the Chicago city limits. The dataset is owned and updated daily by the Chicago Police Department (CPD) and covers traffic crashes from 2013 to the present (2022), although the data includes all Chicago police districts only after 2017. There are also a handful of crashes from 2013 and 2014 (less than 10 total) in the dataset, but this written report will only analyze crash data from 2016 to 2021 due to irregularities in the data entry frequencies in prior years and incomplete data for 2022. The information in the table includes crash parameters such as street condition, weather condition, date, injury counts, and crash type, among others, with 49 distinct columns in total. As of May 2022, the entire dataset contained roughly 610,000 crash entries, with 560,000 from 2016 to 2021.

From 2016 onwards, there have been an average of 94,280 crashes per year, which averages to 258 crashes per day, 11 per hour, and one crash every six minutes. Surprisingly, Chicago does not typically fall on lists of the deadliest cities in the US for drivers, despite its high car accident rate.

A past analytical review of this dataset by legal blog Dopplr covered crashes from 2016 to 2020 and found that the most traffic crashes occurred during October. While October is not typically a wet month, there was more autumn rainfall during this five-year time period compared to the historical average.[1]

The weekday with the most crashes is Friday (Figure 1) with 92,000 crashes total between 2016-2021. Friday is the end of the workweek, the day many people leave for weekend travel, and a common day for dining out and nightlife. Likewise, the most common hours for crashes were between 2:00 pm and 5:00 pm, which are common commute hours.

From the coordinate pairs provided in the dataset, we can see that the most common crash site is around O'Hare Airport, with nearly 800 crashes recorded there between 2016 and 2021. This is followed by Lake Shore Drive on a stretch just north of Navy Pier, with 540 crashes, which is often a congested area, and there are just a few lanes in each direction. Oddly enough, though, the third most frequent coordinate pair belongs to an intersection on 79th street in Southside Chicago, with 400 crashes. Although this location is not downtown or at a heavily trafficked

[1] Dopplrstaff. "When Do the Most Car Accidents Occur in Chicago?: 2016 - 2020." Dopplr Legal News and Information, Dopplr Legal News and Information, 9 Mar. 2022, https://www.dopplr.com/when-do-most-car-accidents-occur-in-chicago-2016-2020/.

airport, this intersection contains several poorly marked turn lanes, making it difficult for drivers to navigate safely.[2]

The day with the most crashes in the dataset was January 12, 2019, with a total of 583 crashes. On that day, the city observed over 6 inches of snowfall, and there were reports of widespread ice and fog. Additionally, this was the year of Chicago's most recent polar vortex, which spanned January and February of 2019 and resulted in record low temperatures in the city. On average, there are 53 crashes per day at times when there was inclement weather (not clear weather). People also tend to drive less on days with bad weather, leading to less crashes.

To see if weather influences car crash frequencies, we investigated the relationship between traffic crashes and weather conditions. Of the crashes caused by the weather (crash cause is noted by the CPD in the records), 42.1% were caused by snow, and 36.1% were caused by rain. The third most crashes were caused by "clear" weather, which still includes slick or icy roads; in the same subset of the data, the road conditions included wet, ice, and snow or slush regardless of the current weather (Figure 2). This finding contradicts the Federal Highway Administration, which claims that most weather-related crashes are due to rainy conditions[3].

However, we plotted the injury count percentages of crashes with clear weather vs. poor weather (rain/snow/sleet/etc.) and found that there was very little difference between the two groups. Crashes with no injuries composed roughly the same percentage (~85%) in both the clear and inclement weather groups; crashes with one injury, two injuries, and three injuries followed the same pattern (Figure 3).

Contrary to popular opinion, our analysis of the data showed that there is a weak relationship between inclement weather and the severity of car crashes. In general, too, only 1.5% of the crashes were caused by weather at all. To further substantiate this observation, we examined the relationship between weather (rain in particular as well as snow) and injury counts through linear regressions with binary variables, excluding rows that contained null values in injury totals.

As shown in Figure 4, there is not a significant difference between the severity of crashes that were caused by clear versus rainy conditions. The slope value is less than 0.05, and the intercept was approximately 0.2, indicating that the difference in the averages between total injuries caused by rain was less than 1 and the average number of injuries on a clear day was effectively 0. Figure 5 compares clear conditions to snow conditions, another prevalent and hazardous weather pattern in Chicago. This regression offered similar results, with the slope value being approximately -0.03, indicating that the difference between total injuries caused by snow was

[2] Anna Poon drives past this intersection once weekly for a UChicago RSO and can confirm that the turn lanes are hazardous.

[3] FHWA. "How Do Weather Events Impact Roads? - FHWA Road Weather Management." FHWA Office of Operations, 20 February 2020, https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm. Accessed 12 May 2022.

also less than 1. Overall, it appears that severe weather conditions such as rain and snow are not correlated with the total number of injuries from a crash. The regression graphs (Figures 4 and 5) demonstrate that there's a greater range in total injury values for clear weather conditions than either snow or rainy weather conditions. This is likely due to random chance and anomalies in the data since there are approximately 9x more clear weather crashes than rainy crashes, and approximately 22x more than snowy weather crashes. We determined that the regressions were not particularly informative due to the large number of crashes that occurred with no injuries. As crashes account for accidents such as rear ends, fender benders, or drive-bys, not all crashes can be considered severe. Thus, we conducted further regression analyses, excluding the crashes that did not result in injuries. Though we expected this approach to yield more significant results and slope values, this was not the case. The slope value decreased to a value of 0.002, confirming that the differences of the average between clear weather crashes and rainy crashes were negligible. For total injuries caused by clear weather versus snow, the slope became slightly more negative for a value of -0.45. However, these values still aren't significant, as they are less than 1.

To test if weather, crash cause, and the number of injuries could act as a predictor of damage costs, we created a k-NN algorithm to classify costs into three categories: less than $500, between $500 and $1500, and above $1500. The crash causes were sorted into external factors vs. driver carelessness and error; injuries were sorted into zero vs. non-zero totals; the weather was sorted into clear vs. inclement weather. However, the classification had an accuracy of between 20-30%, which is not any better than a guess between the three classes. Despite the three factors being logical predictors of damage costs (the severity of the crash for insurance purposes, for example), they returned a poor accuracy.

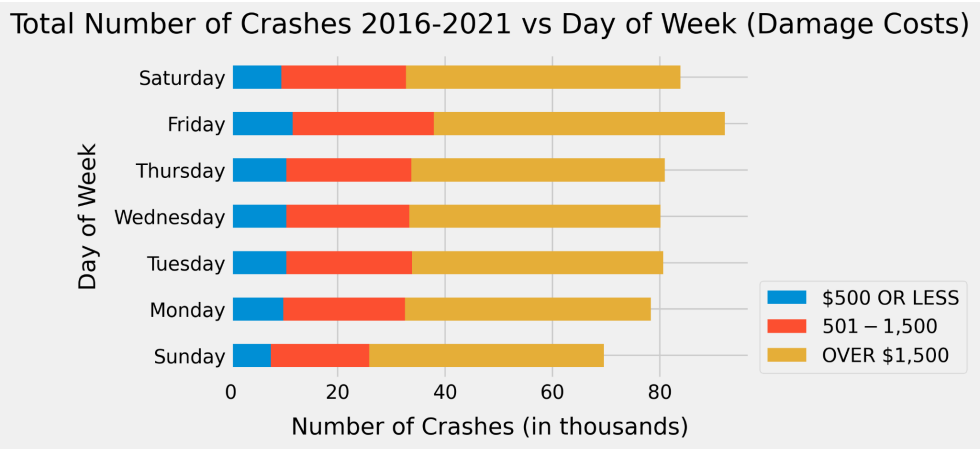

Figure 1: Bar graph showing car crash totals by day of the week from 2016 to 2021 with damage cost categories

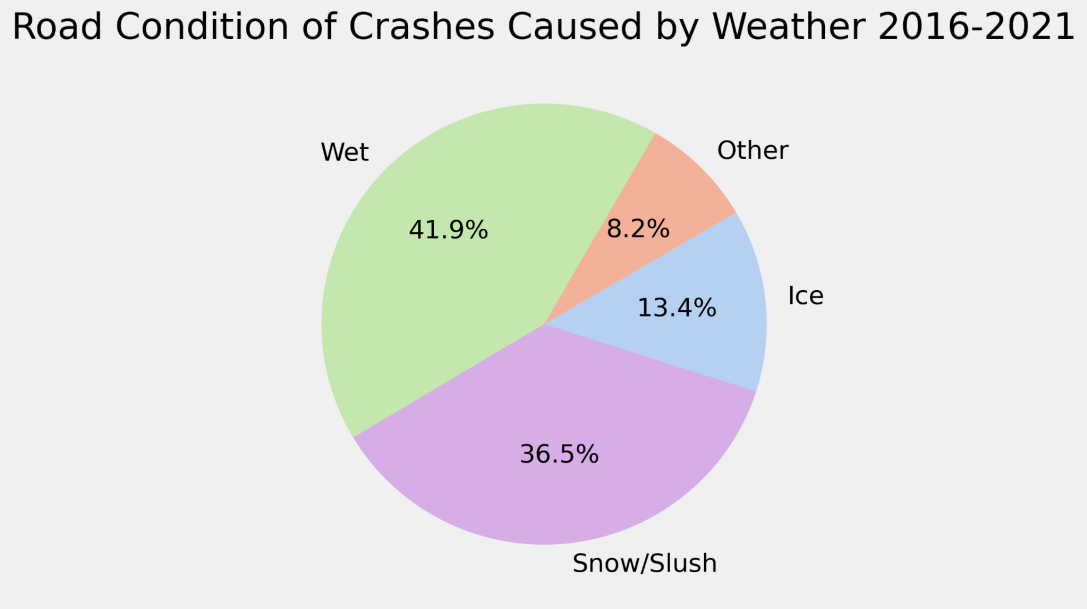

Figure 2: Pie chart of the road condition at the crash site for crashes caused by weather (as noted by the CPD in the dataset)

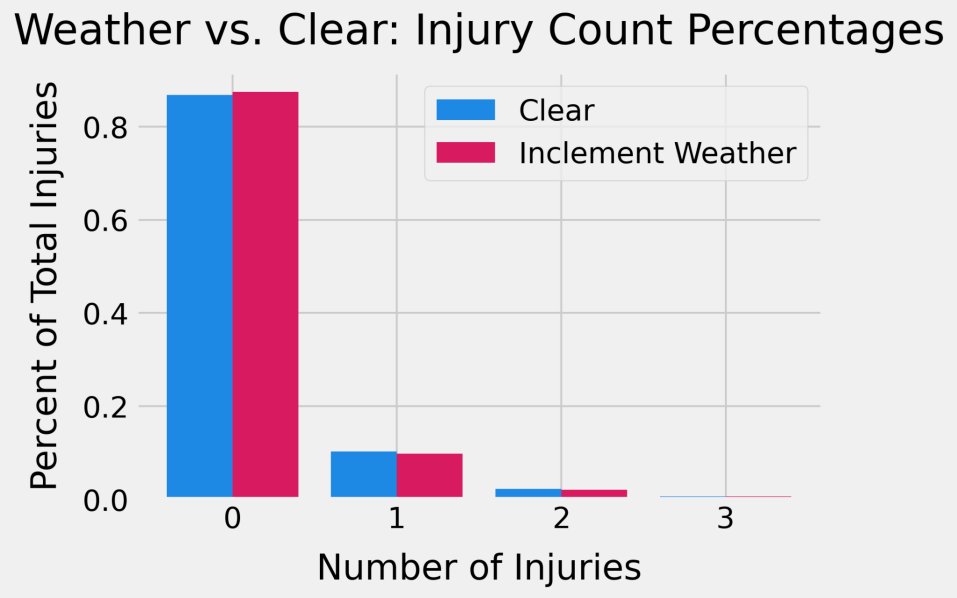

Figure 3: Distribution of injury counts (as percentages of total crashes in each subset) in crashes with clear weather vs. inclement (rain/snow/etc.) weather

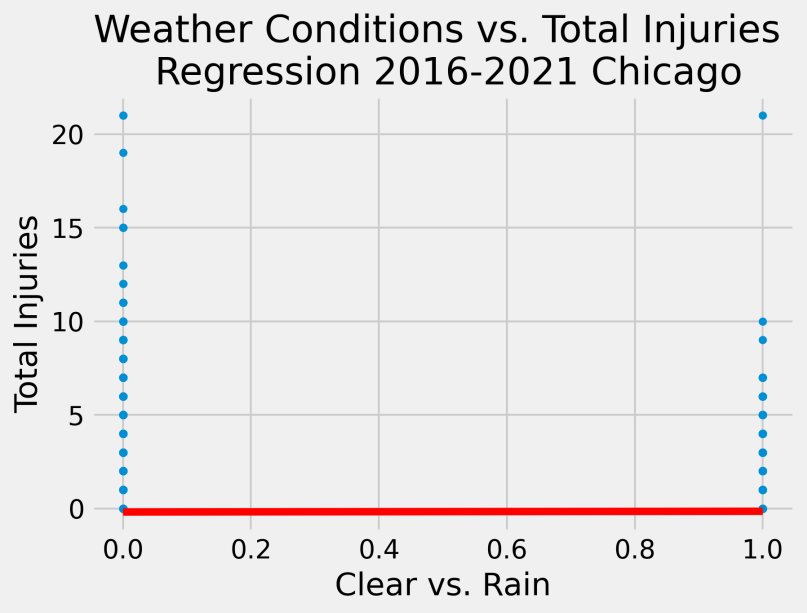

Figure 4: Regression of total number of injuries that were caused by clear vs. rainy conditions

---

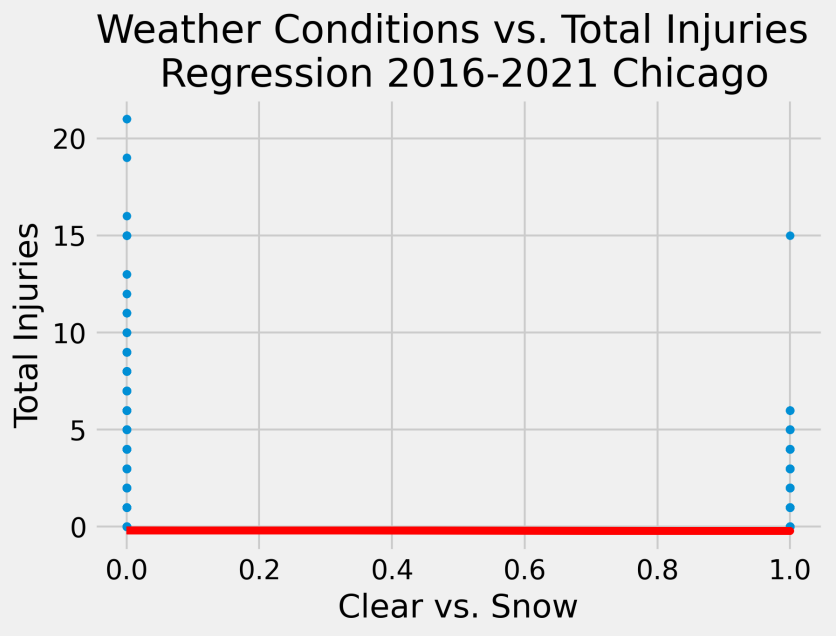

Figure 5: Regression of total number of injuries that were caused by clear vs. snow conditions

Citations:

Author/Publisher: data.cityofchicago.org
Data Owner: Jonathan Levy
Data Provided By: Chicago Police Department
Dataset Name: Traffic Crashes - Crashes
Date Created: October 19, 2017, includes [3/3/2013-4/27/2022]
URL: https://data.cityofchicago.org/Transportation/Traffic-Crashes-Crashes/85ca-t3if

1. US Department of Commerce, NOAA. "January 12, 2019: Widespread Snowfall Event Returns Area to Winter Weather." National Weather Service, NOAA's National Weather Service, 14 Jan. 2019, https://www.weather.gov/lot/12Jan2019_snow.
2. US Department of Commerce, NOAA. "Dangerous Cold of Jan 29-31, 2019." National Weather Service, NOAA's National Weather Service, 1 Feb. 2019, https://www.weather.gov/arx/jan3019.
3. Gilson, D. "10 Deadliest Cities for Drivers (+Insurance Implications): US Insurance Agents." US Insurance Agents | US Insurance Agents Works Hard to Provide Our Users with a Fast and Simple Way to Get and Compare Insurance Rates for Multiple Lines of Insurance., 27 Apr. 2022, https://www.usinsuranceagents.com/deadliest-cities-for-drivers/.
4. Dopplrstaff. "When Do the Most Car Accidents Occur in Chicago?: 2016 - 2020." Dopplr Legal News and Information, Dopplr Legal News and Information, 9 Mar. 2022, https://www.dopplr.com/when-do-most-car-accidents-occur-in-chicago-2016-2020/.
5. FHWA. "How Do Weather Events Impact Roads? - FHWA Road Weather Management." FHWA Office of Operations, 20 February 2020, https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm. Accessed 12 May 2022.

Author Contributions:

Anna:

- Figures 1, 2, 3
- Written report until mention of Figure 3 (page 1, part of page 2)
- k-NN classification exploration

Kaya:

- Figures 4, 5
- Written report concerning regression analysis and some weather information sources