STAT 480 Project

Christopher Cebra and Farid Saud

December 13, 2024

1 Introduction and Data Exploration

1.1 Problem

Traffic crashes are a critical issue facing American cities and society today, being one of the leading causes of fatalities and serious injuries nationwide, especially among people under 50. The question of how to decrease the number of crashes and to decrease their severity have been important challenges for many years, and improvements to roads and vehicles have been suggested for years. Some of these recent improvements, especially focusing on protecting vulnerable pedestrians and cyclists, are protected bike lanes and left-turn traffic calming, a set of speed bumps and bollards that affect where a car can turn left[Chi23].

The City of Chicago Data Portal collates a list of nearly 900,000 traffic crashes that took place since the beginning of 2015[Cit]. Based on this data, we can see that these crashes resulted in over 170,000 injuries and 1,055 fatalities in just the last ten years. This dataset includes one line for every crash with a filed crash report, with 48 columns of information about each crash. A list of all such variables can be found in Table 1. For the variables that are factorized, the factor levels can be found in Table 2. Also, note that this dataset updates daily with new crash reports. All models in this report are carried out on a version of this dataset downloaded on November 5th, 2024.

1.2 Dataset

Table 1: Description of all variables in the City of Chicago crashes dataset, .

Variable Name	Count	Type	Explanation
CRASH_RECORD_ID	889,930	String	ID of filed crash report.
CRASH_DATE_EST_I	65,794	Boolean	Whether the date of the crash was estimated.
CRASH_DATE	889,930	Time	The date and time of the crash.
POSTED_SPEED_LIMIT	889,930	Numeric	The posted speed limit on the road of the crash site.
TRAFFIC_CONTROL _DEVICE	889,930	Factor	Whether a traffic control device was present.
DEVICE_CONDITION	889,930	Factor	If the traffic control device was correctly functioning.
WEATHER_CONDITION	889,930	Factor	What type of weather was taking place at the crash time.
LIGHTING_CONDITION	889,930	Factor	What light level was present at the time of the crash.
FIRST_CRASH_TYPE	889,930	Factor	Classification of crash (rear-end, head-on, etc.).
TRAFFICWAY_TYPE	889,930	Factor	Classification of road on which the crash took place.
LANE_CNT	199,020	Numeric	Number of lanes on the trafficway on which the crash occurred.
ALIGNMENT	889,930	Factor	Type of road alignment on which the crash occurred.
ROADWAY_SURFACE _COND	889,930	Factor	State of the road surface at the time of the crash.
ROAD_DEFECT	889,930	Factor	If there are any defects in the road surface on which the crash occurred.
REPORT_TYPE	862,434	Factor	What type of police report was filed after the crash.
CRASH_TYPE	889,930	Factor	Whether the crash resulted in injuries or a town-away car.
INTERSECTION _RELATED_I	204,225	Boolean	Whether the crash is related to an intersection.

Continued on next page

Table 1 – continued from previous page

Variable Name	Count	Type	Explanation
NOT_RIGHT_OF_WAY_I	40,582	Boolean	Whether the crash was caused by someone not having right-of-way.
HIT_AND_RUN_I	279,117	Boolean	Whether one driver fled the scene of the crash without filing a report.
DAMAGE	889,930	Factor	Three-step factorized factor of crash damage.
DATE_POLICE_NOTIFIED	889,930	Time	Time that the police were notified of the crash.
PRIM_CONTRIBUTORY _CAUSE	889,930	Factor	Assessed primary cause of the crash.
SEC_CONTRIBUTORY _CAUSE	889,930	Factor	If applicable, assessed secondary cause of the crash.
STREET_NO	889,930	Numeric	House number of street address of crash location.
STREET_DIRECTION	889,930	String	Street direction of street address of crash location.
STREET_NAME	889,926	String	Name of street address of crash location.
BEAT_OF_OCCURRENCE	889,925	Factor	Police beat (district) of crash location.
PHOTOS_TAKEN_I	12,147	Boolean	Whether photos of the crash location were taken.
STATEMENTS_TAKEN_I	20,452	Boolean	Whether statements about the crash were taken.
DOORING_I	2,825	Boolean	Whether crash involved a motor vehicle occupant opening a door into the travel path of a bicyclist, causing a crash.
WORK_ZONE_I	4,996	Boolean	Whether the crash took place in a road work zone.
WORK_ZONE_TYPE	3,860	Factor	What type of work zone the crash took place in.
WORKERS_PRESENT_I	1,286	Boolean	Whether workers were present in the work zone.
NUM_UNITS	889,930	Numeric	How many units (vehicles, bicycles, pedestrians, etc.) involved in the crash.
MOST_SEVERE_INJURY	887,952	Factor	Most severe injury of participants in the crash (or no injury).
INJURIES_TOTAL	887,966	Numeric	Total number of injured individuals in the crash (total is 172,271).
INJURIES_FATAL	887,966	Numeric	Total number of fatally injured individuals in the crash (total is 1,055).
INJURIES _INCAPACITATING	887,966	Numeric	Total number of incapacitated individuals in the crash (total is 17,559).
INJURIES_NON _INCAPACITATING	887,966	Numeric	Total number of nonincapacitated but injured individuals in the crash (total is 96,630).
INJURIES_REPORTED _NOT_EVIDENT	887,966	Numeric	Total number of individuals reporting injuries but not evident to crash reporters.
INJURIES_NO_INDICATION	887,966	Numeric	Total number of individuals involved in crashes but uninjured.
INJURIES_UNKNOWN	887,966	Numeric	Unknown injury status.
CRASH_HOUR	889,930	Numeric	Hour that the crash occurred.
CRASH_DAY	889,930	Numeric	Day that the crash occurred.
CRASH_MONTH	889,930	Numeric	Month that the crash occurred.
LATITUDE	883,544	Numeric	Latitude of precise crash location.
LONGITUDE	883,544	Numeric	Longitude of precise crash location.
LOCATION	883,544	Spatial	Spatial point variable of crash latitude and longitude.

Table 2: List of levels for the factorized variables in the Chicago crashes dataset.

Variable Name	**Factors**
TRAFFIC_CONTROL _DEVICE	Bicycle Crossing Sign, Delineators, Flashing Control Signal, Lane Use Marking, No Controls, No Passing, Other, Other Railroad Crossing, Other Reg. Sign, Other Warning Sign, Pedestrian Crossing Sign, Police/Flagman, Railroad Crossing Gate, RR Crossing Sign, School Zone, Stop Sign/Flasher, Traffic Signal, Unknown, Yield
DEVICE_CONDITION	Functioning Improperly, Functioning Properly, Missing, No Controls, Not Functioning, Other, Unknown, Worn Reflective Material
WEATHER_CONDITION	Blowing Sand, Soil, Dirt, Blowing Snow, Clear, Cloudy/Overcast, Fog/Smoke/Haze, Freezing Rain/Drizzle, Other, Rain, Severe Cross Wind Gate, Sleet/Hail, Snow, Unknown
LIGHTING_CONDITION	Darkness, Darkness, Lighted Road, Dawn, Daylight, Dusk, Unknown
FIRST_CRASH_TYPE	Angle, Animal, Fixed Object, Head On, Other Noncollision, Other Object, Overturned, Parked Motor Vehicle, Pedalcyclist, Pedestrian, Rear End, Rear to Front, Rear to Side, Sideswipe Opposite Direction, Sideswipe Same Direction, Train, Turning
TRAFFICWAY_TYPE	Alley, Center Turn Lane, Divided - w/Median (Not Raised), Divided - w/Median Barrier, Driveway, Five Point, or More, Four Way, L-Intersection, Not Divided, Not Reported, One-Way, Other, Parking Lot, Ramp, Roundabout, T-Intersection, Traffic Route, Unknown, Unknown Intersection Type, Y-Intersection
ALIGNMENT	Curve on Grade, Curve on Hillcrest, Curve, Level, Straight and Level, Straight on Grade, Straight on Hillcrest
ROADWAY_SURFACE _COND	Dry, Ice, Other, Sand, Mud, Dirt, Snow or Slush, Unknown, Wet
ROAD_DEFECT	Debris on Roadway, No Defects, Other, Rut, Holes, Shoulder Defect, Unknown, Worn Surface
REPORT_TYPE	Amended, Not on Scene (Desk Report), On Scene
CRASH_TYPE	Injury and/or Tow Due to Crash, No Injury/Drive Away
DAMAGE	$500 or Less, $501-$1,500, Over $1,500
PRIM_CONTRIBUTORY _CAUSE	Animal, Bicycle Advancing Legally on Red Light, Cell Phone Use, Disregarding Other Traffic Signs, Disregarding Road Markings, Disregarding Stop Signs, Disregarding Traffic Signals, Disregarding Yield Sign, Distraction From Inside Vehicle, Distraction From Outside Vehicle, Distraction - Other Electronic Device, Driving in Wrong Side/Wrong Way, Driving Skills/Knowledge/Equipment, Equipment-Vehicle Condition, Evasive Action due to Animal, Object, Nonmotorist, Exceeding Authorized Speed Limit, Exceeding Safe Speed for Conditions, Failing to Reduce Speed to Avoid Crash, Failing to Yield Right-of-Way, Following Too Closely, Had Been Drinking (no arrest), Improper Backing, Improper Lane Usage, Imroper Overtaking/Passing, Improper Turning/No Signal, Motorcycle Advancing Legally on Red Light, Not Applicable, Obstructed Crosswalks, Operating Vehicle in Erratic, Reckless Manner, Passing Stopped School Bus, Physical Condition of Driver, Related to Bus Stop, Road Construction/Maintenance, Road Engineering/Surface/Marking Defects, Texting, Turning Right on Red, Unable to Determine, Under the Influence of Alcohol/Drugs (with Arrest), Vision Obscured (Signs, Tree Limbs, Buildings, etc.), Weather
SEC_CONTRIBUTORY _CAUSE	(same as above)

1.3 Preliminary Data Analysis and Research Question

In total, our dataset of 889,930 rows and 48 columns summed to 481.7 MB, clearly meeting the threshold for "big data". When we converted all of the factors into a model matrix by factor level, we ended up with over 260 total explanatory variables, and over 1.8 GB. Clearly, some sort of variable selection criterions would have to be applied.

Each row of the dataset corresponds to an individual crash. Since we do not know the overall prevalence of every condition in society, it is not possible to calculate the rate that crashes occurred. Therefore, we will focus on the severity of crashes that did occur. Will will focus on the variables CRASH_TYPE and MOST_SEVERE_INJURY as our potential response variables.

Given these conditions and response variables, we ask the following three **research questions**:

1. **Are there any explanatory variables which are predictive of crash severity?**
2. **Are there variables which are causal with crash severity?**
3. **Can we quantify the performance of street design improvements (left-turn traffic calming and bike lanes) as an experiment to analyze their performance?**

We plan to use LASSO for variable selection, and double LASSO to test for causality. We want to maximize efficiency, so we carry out LASSO with AICc as the model selection criteria because of its relatively light computations. We also performed 5-fold cross-validation LASSO for a "more stable" solution, leveraging multiple cores for parallel computation. We used both R and Python, with sklearn.linear_model.LogisticRegressionCV() in Python for cross-validation-driven LASSO and gamlr() in R for AICc-based selection to run our LASSO models.

1.4 Data Cleaning

We first carry out some preliminary data processing and cleaning to address the first research question.

1. The dataset does not have a lot of missing data except in certain variables, like the Boolean variables, where it is blank if the condition is not true. We will have to fill all of these blank observations with "N" for each Boolean variable. For missing values in REPORT_TYPE and WORK_ZONE_TYPE, we filled them with "UNKNOWN". LANE_CNT contained a lot of missing data, including several rows with over 100 listed lanes, we factorized it as NARROW if the number of lanes was 2 or less, WIDE if it was 3-4, HIGHWAY for anything above this, and NOT_APPLICABLE otherwise.
2. We wanted to engineer a few new variables that we thought could be explanatory. After converting the CRASH_DATE and DATE_POLICE_NOTIFIED variables to datetime format, we calculated the difference between the two in hours as Report_vs_Police_Notified. We also extracted the crash year, which we called Crash_Year.
3. Since we aimed to predict the severity of crashes, we removed variables that took place after the crash: PHOTOS_TAKEN_I, STATEMENTS_TAKEN_I, and locational variables with too many factors to be analyzable: CRASH_RECORD_ID, STREET_NO, STREET_DIRECTION, STREET_NAME, BEAT_OF_OCCURRENCE, LOCATION, LATITUDE and LONGITUDE. As a location variable, we took the police beats variable and factorized it by Chicago's 25 police districts. For a map of the police beats and their police districts, see [Chi].
4. Every boolean and categorical variables were transformed to factor type variables. For a reference, the baseline level for every categorical variable is presented in figure For all categorical variables, the reference levels are summarized in Figure 1.

Variable	Reference Level	Variable	Reference Level
CRASH_DATE_EST_I	N	HIT_AND_RUN_I	N
TRAFFIC_CONTROL_DEV	Bicycle Crossing Sign	DAMAGE	$500 or Less
DEVICE_CONDITION	Functioning Improperly	PRIM_CONTRIBUTORY_CAUSE	Animal
WEATHER_CONDITION	Sand, Soil, Dirt	SEC_CONTRIBUTORY_CAUSE	Animal
LIGHTING_CONDITION	Darkness	DOORING_I	N
FIRST_CRASH_TYPE	Angle	WORK_ZONE_I	N
TRAFFICWAY_TYPE	Alley	WORK_ZONE_TYPE	Construction
LANE_CNT	Highway	WORKERS_PRESENT_I	N
ALIGNMENT	Curve on Grade	CRASH_HOUR	0
ROADWAY_SURFACE_COND	Dry	CRASH_DAY_OF_WEEK	1
ROAD_DEFECT	Debris on Roadway	CRASH_MONTH	1
REPORT_TYPE	Amended	Crash_Year	2013
INTERSECTION_RELATED_I	N	Police_district	District 01
NOT_RIGHT_OF_WAY_I	N		

Figure 1: Reference Levels for Categorical Variables

To deal with the size of our data, we split the intial and processed, clean datasets into 5 subsets (or "folds") so that they would fit in Github, and merged them together in R and Python afterwards.

2 Research Question 1

2.1 Model Building

We focused on LASSO models for the first research question, to identify the significant variables towards crash severity. We focused on two different response variables: (1) CRASH_TYPE and (2) MOST_SEVERE_INJURY. CRASH_TYPE has only two levels: (i) No Injury/Drive Away and (ii) Injury And/Or Tow Due to Crash. This allows for a good mark for crash severity. For MOST_SEVERE_INJURY, there are five levels: (i) No Indication of Injury, (ii) Reported but Not Evident, (iii) Non-incapacitating Injury, (iv) Incapacitating Injury, and (v)Fatal.

In Figure 2 we can see the distribution of factor levels for both variables. The Crash Type variable is more balanced, and will be easier to perform regression on, while the Most Severe Injury, for which we'll perform multinomial regression on, has a lot more imbalance. Most crashes show no indication of injury, while only 1.7% of crashes resulted in an incapacitating, and 0.1% in a fatal, injury. The more balanced variables allow for a good sample size of each level, making fitting the LASSO model easier.

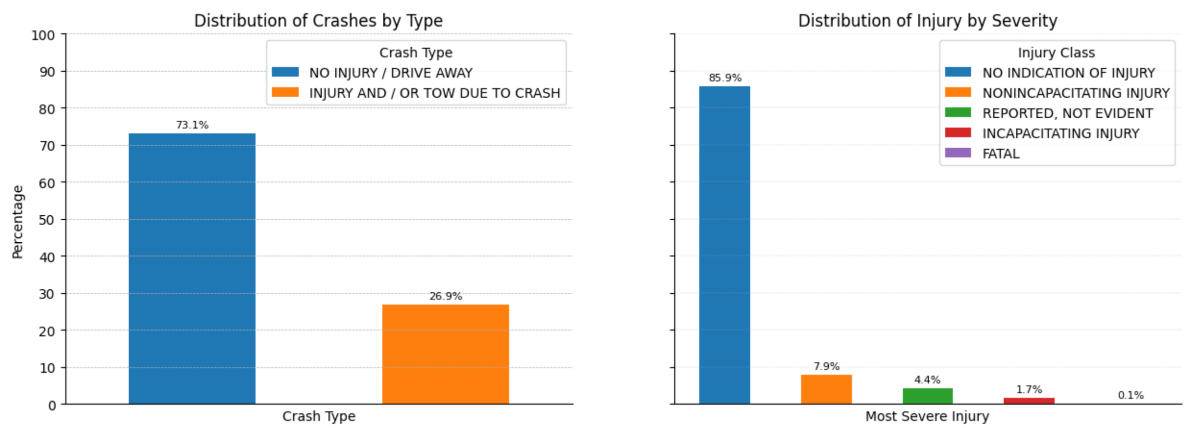

Figure 2: Distribution of factor levels of Response variables: Crash Type and Most Severe Injury.

For fitting LASSO models, we standardized all variables so that the mean and variance of each column are 0 and 1. R's gamlr() function allowed us to perform this transformation by setting 'standardize=TRUE', whereas for Python's sklearn.linear_model. LogisticRegressionCV() we used sklearn.preprocessing.StandardScaler to .fit_ transform(X).

From there, we fitted binomial/multinomial LASSO models. We ran our code on a High-Performance Computing cluster with the interlagos node, quad-socket AMD Interlagos with 64 cores per node and 512 GB of memory.

2.2 Models' Results

We started with AICc gamlr() for the binary response Crash Type with default parameters. The initial results had selected the smallest lambda in the standard range, which pointed to a potential corner solution. This behavior suggested that the default lambda sequence might not be sufficiently broad to capture the optimal regularization parameter. To address this, we expanded the lambda range to include smaller values by adjusting the lambda.min.ratio, but the model continued to converge at the smallest lambda, eventually failing to zero out variables at all.

To balance computational efficiency and the risk of over-fitting, we avoided excessively expanding the lambda range beyond what we thought was reasonable based on exploratory tests. Continuing to test smaller and smaller lambda values could result in overly complex models that fail to generalize well, as the penalty becomes too weak to eliminate insignificant variables. For this, we set lambda.min.ratio = 1e-4.The top and bottom 10 coefficients and the corresponding variables are presented for the default and our selected lambda ranges are presented in 3 and 4:

Variable	Coefficient
First Crash Year: 2015	1.245
First Crash Year: 2016	1.153
First Crash Type: Sideswipe Same Dir.	1.151
First Crash Type: Rear to Rear	1.047
Primary Cause: Improper Backing	1.027
First Crash Type: Rear to Front	0.902
Trafficway Type: Parking Lot	0.827
First Crash Type: Rear End	0.692
First Crash Type: Parked Motor Vehicle	0.573
Trafficway Type: Driveway	0.548

(a) Top 10 coefficients.

Variable	Coefficient
First Crash Type: Overturned	-3.411
First Crash Type: Pedestrian	-2.957
First Crash Type: Fixed Object	-1.872
First Crash Type: Cyclist	-1.754
First Crash Type: Train	-1.681
Primary Cause: DUI	-1.563
First Crash Type: Other Noncollision	-1.413
Primary Cause: Driver Condition	-1.378
First Crash Type: Other Object	-1.230
Primary Cause: Vehicle Condition	-1.113

(b) Bottom 10 coefficients.

Figure 3: Coefficients for binomial AICc - LASSO model, default lambda.min.ratio = 0.01.

Variable	Coefficient
First Crash Type: Rear to Rear	1.738
First Crash Year: 2015	1.509
First Crash Year: 2016	1.320
First Crash Type: Sideswipe Same Dir.	1.236
Primary Cause: Improper Backing	1.209
Primary Cause: Bicycle Advancing Legal	1.158
First Crash Type: Rear to Front	1.082
First Crash Type: Rear End	0.780
First Crash Type: Parked Motor Vehicle	0.690
First Crash Type: Rear to Side	0.677

(a) Top 10 coefficients.

Variable	Coefficient
First Crash Type: Overturned	-3.863
First Crash Type: Pedestrian	-3.025
First Crash Type: Train	-2.640
First Crash Type: Fixed Object	-1.856
First Crash Type: Cyclist	-1.771
First Crash Type: Other Noncollision	-1.517
Crash Year: 2014	-1.469
Primary Cause: DUI	-1.455
Primary Cause: Driver Condition	-1.284
First Crash Type: Other Object	-1.261

(b) Bottom 10 coefficients.

Figure 4: Coefficients for binomial AICc - LASSO model, extended lambda.min.ratio = 1e-4.

We decided to explore cross-validation as an alternative selection method, as CV allows for potentially identifying a more balanced lambda due to the averaging of the folds. We transitioned to Python's sklearn.linear_model. LogisticRegressionCV() since it offered the additional benefit of automatically handling binomial logistic regression for binary responses as well as multinomial logistic regression for multi-level responses. To mitigate the computational overhead of running cross-validation, we set the number of folds to k=5, and leveraged multi-core parallel computation with the argument n_jobs=5.

While for our multi-level response Most Severe injury, variable selection was performed successfully, cross-validation LASSO for the binary response Crash Type once again did not "zero out" any variable. The top and bottom 10 coefficients and variables are presented for these models in figure 5 for Crash Type, and in figures 6 and 7 for Most Severe Injury:

Variable	Coefficient
Report Type: Not on Scene (Desk)	0.866
First Crash Type: Sideswipe Same Dir.	0.461
First Crash Type: Parked Motor Vehicle	0.409
First Crash Type: Rear End	0.264
Primary Cause: Improper Backing	0.232
Hit and Run: Yes	0.162
First Crash Type: Turning	0.135
First Crash Hour: 16	0.112
First Crash Type: Rear to Front	0.111
Crash Hour: 17	0.110

(a) Top 10 coefficients.

Variable	Coefficient
First Crash Type: Pedestrian	-0.415
Number of Units	-0.401
Trafficway Type: Not Divided	-0.379
Trafficway Type: Divided (No Median)	-0.313
First Crash Type: Fixed Object	-0.240
Report Type: On Scene	-0.240
Trafficway Type: Median Barrier	-0.238
Trafficway Type: Four-Way	-0.205
First Crash Type: Cyclist	-0.191
Trafficway Type: One-Way	-0.173

(b) Bottom 10 coefficients.

Figure 5: Coefficients for binomial CV-LASSO model.

Variable	Coefficient
Report Type: Not on Scene (Desk)	0.799
First Crash Type: Parked Motor Vehicle	0.561
First Crash Type: Sideswipe Same Dir.	0.282
Hit and Run: Yes	0.149
Primary Cause: Improper Backing	0.130
Trafficway Type: Parking Lot	0.093
Trafficway Type: One-Way	0.057
Lighting: Unknown	0.047
First Crash Type: Turning	0.039
First Crash Type: Rear to Front	0.029

(a) Top 10 coefficients.

Variable	Coefficient
First Crash Type: Pedestrian	-0.620
Damage: Over $1,500	-0.353
First Crash Type: Cyclist	-0.291
Number of Units	-0.192
Intersection Related: Yes	-0.136
Speed Limit	-0.108
Primary Cause: Driver Condition	-0.103
Police District: 11	-0.088
Primary Cause: Traffic Signal Violation	-0.078
Police District: 06	-0.072

(b) Bottom 10 coefficients.

Figure 6: Coefficients for multinomial response CV-LASSO model (Level: No Indication of Injury).

Variable	Coefficient
First Crash Type: Cyclist	0.087
First Crash Type: Pedestrian	0.067
Primary Cause: Driver Condition	0.032
Damage: Over $1,500	0.031
First Crash Type: Fixed Object	0.022
First Crash Type: Head On	0.003
Primary Cause: DUI	0.003
Secondary Cause: DUI	0.000
Secondary Cause: Unable to Determine	0.000
Secondary Cause: Right Turn on Red	0.000

(a) Top 10 coefficients.

Variable	Coefficient
Report Type: Not on Scene (Desk Report)	-0.106
First Crash Type: Rear End	-0.015
Secondary Cause: Motorcycle Red Light Violation	0.000
Secondary Cause: Not Applicable	0.000
Secondary Cause: Obstructed Crosswalks	0.000
Secondary Cause: Reckless Driving	0.000
Secondary Cause: Passing School Bus	0.000
Secondary Cause: Driver Condition	0.000
Secondary Cause: Related to Bus Stop	0.000
Secondary Cause: Road Maintenance	0.000

(b) Bottom 10 coefficients.

Figure 7: Coefficients for multinomial response CV-LASSO model (Level: Incapacitating Injury).

At this point, given that both AICc and CV-based LASSO models were yielding similar results in terms of (not) regularizing the variables, we decided to fall back to R's gamlr(), in particular for the 'free = ...' argument, which allows to un-penalize a specific variable. This tool allowed us to perform Double LASSO models for testing Causality.

2.3 Predictive Variables Analysis

The selected model for the Causal Analysis is the AICc-based LASSO model for our binary response Crash Type, with the extended lambda range (variables and coefficients in 4). This model zero-ed out only 2 level variables: Secondary Cause: PASSING STOPPED SCHOOL BUS and Crash Year: 2020. From these results, these 2 variables are not predictive of the crash type. Apart from this, every variable seemed to be predictive of crash type. From 4:

- The variables with positive coefficients are associated with No Injury/Drive Away (less severe crashes). The odds of a crash being less severe would increase by e^{coef}, compared to the reference level. This is, for example, the odds of a crash being less severe increase by $e^{1.738} = 5.686$ (an increase of 468.6%) if the First Crash Type is a "Rear to Rear", as compared to the First Crash Type being "Angle" (reference level). Similarly, if the First Crash Type is a "Rear to Side", the odds of the crash being less severe increase by $e^{0.677} = 1.968$ (an increase of 96.8%), as compared to "Angle".
- In contrast, the variables with negative estimated coefficients are associated with more severe crashes: Injury and/or Tow due to crash. The odds of a crash being less severe decrease by $1 - e^{\text{coef}}$. For instance, when the First Crash Type is "Pedestrian", the odds of a crash being less severe decrease by $1 - e^{-3.025} = 0.952$ (a 95.2% decrease). If the First Crash Type is "Train", the odds decrease by $1 - e^{-2.640} = 0.929$ (92.9% lower odds). Similarly, if the Primary Cause is "DUI", the odds decrease by $1 - e^{-1.455} = 0.767$ (76.7% decrease).

We believe that these results make sense as crash types such as "Rear to Rear" or "Rear to Side" typically involve low-impact collisions, leading to less severe crashes, while crash types like "Pedestrian" or "Train" often involve vulnerable road users or high-impact, resulting in more severe incidents. Similarly, behavioral factors such as "DUI" are well-documented contributors to severe crashes, aligning with the negative coefficients observed in the model.

3 Research Question 2

To address research question 2, we tested for causality using the double LASSO method. Building on the considerations from the earlier "naive LASSO models," we fitted two consecutive LASSO models for each non-zero variable selected by our AICc model with lambda.min.ratio = 1e-4. Each non-zero variable, referred to as a "treatment variable" (d), was first modeled using all other variables to predict the treatment. With this prediction *d_hat*, we fitted a second model where the predicted treatment is part of the model as an unpenalized variable to estimate its causal effect.

From the selected initial LASSO model, 259 variables (not including the intercept) were non-zero. This meant that we had to fit 2 LASSO models 259 times resulting in 518 total LASSO fits. Given the computational intensity of this task, we used a parallel back-end based on the libraries future and future.apply and all the 64 cores of the HPC node. These libraries enabled shared-memory processing, so we do not have to copy the design matrix and response variable to every core for improved computational efficiency.

Variable	**"Naive" LASSO**	**Double LASSO**	$\mathbf{R^2}$
First Crash Type: Rear to Rear	1.7378	0.8939	0.2437
Crash Year: 2015	1.5094	0.0000	0.9980
Crash Year: 2016	1.3201	0.0000	0.9990
First Crash Type: Sideswipe Same Dir.	1.2355	1.0142	0.7589
Primary Cause: Improper Backing	1.2091	0.0000	0.9875
Primary Cause: Bicycle Advancing Legal	1.1579	0.0000	0.7812
First Crash Type: Rear to Front	1.0822	0.8779	0.4438
First Crash Type: Rear End	0.7797	0.7376	0.8098
First Crash Type: Parked Motor Vehicle	0.6896	0.3367	0.8460
Crash Type: Rear to Side	0.6771	0.3790	0.2808

(a) Top 10 coefficients.

Variable	**"Naive" LASSO**	**Double LASSO**	$\mathbf{R^2}$
First Crash Type: Overturned	-3.8625	0.0000	0.8848
First Crash Type: Pedestrian	-3.0245	-2.8933	0.3052
First Crash Type: Train	-2.6397	-1.6815	0.3987
First Crash Type: Fixed Object	-1.8557	-0.3145	0.9223
First Crash Type: Cyclist	-1.7710	-1.6346	0.3425
First Crash Type: Other Noncollision	-1.5169	0.0000	0.8828
Crash Year: 2014	-1.4685	0.0000	0.9929
Primary Cause: DUI	-1.4553	0.0000	0.9245
Primary Cause: Driver Condition	-1.2835	0.0000	0.9350
First Crash Type: Other Object	-1.2612	-0.1509	0.8363

(b) Bottom 10 coefficients.

Figure 8: Naive and double LASSO models coefficients for top and bottom 10 variables

Several variables initially identified as significant in the naive LASSO model were no longer significant after controlling for confounders (green rows in 8). Among these variables, for the ones initially associated with Injuries and Towing due to car crashes (negative coefficients), Crash Year: 2014, Primary Causes: Driver Condition, and DUI, and First Crash Types: Other Non-collision and Overturned, even though associated with more severe crashes, are identified as not causal (given that the coefficients for the double LASSO were 0). Likewise, for the variables initially associated with positive effects in the naive LASSO model, 4 were identified as non-causal. These were Crash Years: 2015, 2016, and Primary Cause: Improper Backing, Bicycle Advancing Legally.

Conversely, the variables that can be assumed to be **causal**, based on the double LASSO results, are in the white rows in 8. First Crash Type: Rear to Rear, Sideswipe Same Direction, Rear to Front, Rear End, Parked Motor Vehicle, and Crash Type: Rear to Side retained non-zero coefficients in the double LASSO model. This indicates causal association with less severe crashes, potentially due to the nature of these crash types as controlled or lower-impact scenarios. First Crash Types: Other Object, Cyclist, Fixed Object, Train, and Pedestrian were also non-zero, which means there is causal association with more severe crashes.

4 Research Question 3

Clearly, as we have seen in Section 2 and 3, two of the most important and causal predictors of serious crashes are crashes with pedestrians and with cyclists. This has been the focal point of numerous traffic safety programs, including the present Vision Zero and Complete Streets, and we wanted to test whether any of the programs that the City of Chicago employed to attempt to reduce these fatalities were successful. One of these are Left-Turn Traffic Calming (LTTC), a series of plastic bollards and speed bumps that aim to improve visibility lanes for left-turning drivers crossing over pedestrian crosswalks, as well as reduce the cars' speed The second was the construction of bike lanes, neighborhood bike greenways, and protected bike lanes around the city.

Our hope in considering these metrics was that these could behave more as an experiment than as an observational study. If we can find a precise date in which these measures were installed, we can find the proportion of crashes originating from before and after they were installed, and use experimental techniques to analyze their differences.

4.1 Left-Turn Traffic Calming (LTTC)

The history of left-turn traffic calming in Chicago originated with a pilot study conducted in Chicago's touristy and busy River North area in 2019, consisting of 5 adjacent intersections along State Street. The results of this study were positive, and the city reported that "Crashes at these intersections reduced by 24% whereas similar intersections on the same corridor did not see any reduction in crashes involving left turning vehicles"[Chi23]. For a picture of what Left-Turn Traffic Calming entails, see [Chi23]. hose in 2022 to install LTTC infrastructure at a further 15 intersections in different neighborhoods around the city. For a full list of these intersections, see:[Chi22b]

We took our original dataset and applied the same variable creations as we did in Section 2. Then, we focused on the five River North intersections, which were State and Kinzie, Hubbard, Illinois, Grand, Ohio and Ontario Streets, so we subsetted by street name to include the cross-streets from 20 West to 20 East (State Street is 0 W/E in the Chicago Street grid) and from 360 to 640 North on State Street (Kinzie is 400 N and Ontario is 628 N). This includes a lead-up to each of the affected intersections. We then also subsetted the crashes to include only INTERSECTION_RELATED = Y. In Figure 9 shows the number of total intersection-related crashes, the number of intersection-related crashes involving pedestrians, and the number of these crashes resulting in injuries or tow-aways. The results from the study area form the first three rows, while as a baseline the second three rows show the same numbers of crashes from the police beats contained within the study area.

We then repeated this study for the intersections in the 2022 LTTC area. The results of this can be seen in Figure 10.

Using these results, we tested the yearly means before and after the treatments were applied. In Figure 9. For the River North area, the 2017-19 mean total crashes was 59, while the 2020-24 mean was 36.8, a 37.6% decrease. For pedestrian-involved crashes, the sample size is much smaller, but the 2017-19 mean was 3.67 and the 2020-24 mean was 3.60. In comparison, for the larger neighborhood area, the 2017-19 mean crash total was 1021 whereas the 2020-24 mean total was 644.2, a 37% decrease, and the pedestrian-involved crashes mean decreased 29.5% from 72.33 to 51. There does not seem to be a notable difference in the amount of the decrease between the LTTC intersections and the other intersections in River North.

For the 2022 intersections in Figure 10, we see a 2018-22 mean total crashes of 86.2 compared to 70.5 after LTTC (an 18.2% decrease). For pedestrian-involved crashes, the 2018-22 mean was 8.6 compared to 5 for 23-24 (41.8% decrease). Across their surrounding neighborhoods, the 2018-22 crash average was 1278.6 and 23-24 was 1089 (a 14.8% decrease), while pedestrian-involved crashes dropped 4.4% from 58.6 to 56.

The results from Figures 9 and 10 are mixed. Although the River North intersections saw a decrease in traffic collisions, the broader neighborhood saw a similar decrease. The situation is stronger for the 2022 survey area, which saw a larger decrease, especially for pedestrian-involved crashes, than their surrounding neighborhoods. However, there are a few factors involving the data and potential confounders to consider. Firstly, the number of crashes in the study areas are not that high, especially for the River North dataset, and if we want to focus only on the ones involving pedestrians, it is negligible. However, LTTC should slow the speed of turning cars in general, which should affect both pedestrian crashes and those involving only cars. Although the number of crashes in the River North study area did decline after 2019 when the LTTC procedures were installed, this is confounded by the overall decrease in crashes in the surrounding police beats. This is likely due to COVID and return-to-office slowdowns, as the River North area has a lot of in-person office work and tourism, both of which likely declined after 2019. Also, LTTC infrastructure, especially its plastic bollards, can break down over time as they collapse when being hit by a car[Chi19].

Secondly, the study intersections for the 2022 installations were selected because they were referred to as "intersections with a history of left-turn crashes." Therefore, if there was some regression to the mean, this might also be classified as improvement due to the LTTC technology. Furthermore, some of the benefits of LTTC, such as "increase in drivers yielding to pedestrians when making left turns" which they state as a major success of the program, cannot be ascertained from the crash data. Focusing only on crashes may make analyzing general road calming behavior harder.

Variable	16	17	18	19	20	21	22	23	24
Crashes	24	55	68	54	36	47	48	27	26
Pedestrian Involved	0	2	3	6	4	3	6	2	3
Injury or Tow	4	6	11	13	10	18	19	12	5
Beat Crashes	352	794	1188	1081	545	700	764	687	525
Beat PI	8	46	72	99	40	57	54	56	48
Beats Injury	34	169	348	319	217	267	267	240	182

Figure 9: LTTC for River North Study Area. Data from the 6 intersections with LTTC installed was compared with their encompassing police beats (1831, 1832, 1834). Blue: The year the LTTC infrastructure was installed at the study intersections. Red: Years since the LTTC infrastructure was installed at the study intersections.

Variable	16	17	18	19	20	21	22	23	24
Crashes	21	52	75	90	87	106	73	79	62
Pedestrian Involved	1	3	12	10	5	10	6	5	5
Injury or Tow	9	23	31	50	42	56	31	38	25
Beat Crashes	411	903	1305	1378	1107	1321	1282	1200	978
Beat PI	12	43	75	77	50	42	49	58	54
Beats Injury	103	366	610	630	574	643	654	598	478

Figure 10: LTTC for 2022 Intersections Study Area. Data from the 15 intersections with LTTC installed was compared with their encompassing police beats (732, 733, 734, 735, 823, 825, 1211, 1212, 1213, 1215, 1221, 1423, 1912, 2512, 2515). Blue: The year the LTTC infrastructure was installed at the study intersections. Red: Years since the LTTC infrastructure was installed at the study intersections.

4.2 Bike lanes

In addition to Left-Turn Traffic Calming, another major infrastructure treatment to improve safety is the construction of bike lanes. As of 2024, the City of Chicago has 497 miles of bikeways, nearly all of which have been constructed since 2010[Nei24].

The city classifies these bikeways into six categories, ranging from "calmest" to least calm (based on amount of interactions with traffic). In order of least calm to most calm, these are Marked Shared Lanes, also called "sharrows", which are roads designated as bikeways where bikes operate on the main road in mixed traffic. Next are bike lanes, with painted delineated lanes with bike markings on the shoulder of the road, but no separation from mixed traffic. After these are buffered bike lanes, which have no physical barriers from mixed traffic but have some setback distance, to reduce the risk of bikes being clipped by moving cars. Next after this are neighborhood greenways. These are essentially just bikes operating in mixed traffic, but the city designates certain calm, low-traffic residential streets with this category to provide calmer options for cyclists. Next are protected bike lanes, which are like buffered bike lanes but have physical barriers (curbs and bollards) to separate them from mixed traffic. Some protected bike lanes are between the sidewalk and streetside parking as well. Finally are off-street trails, such as those through parks, where the travelway is only for bikes and potentially pedestrians. For a picture of each of these kinds of bikeway, see Figure 11.

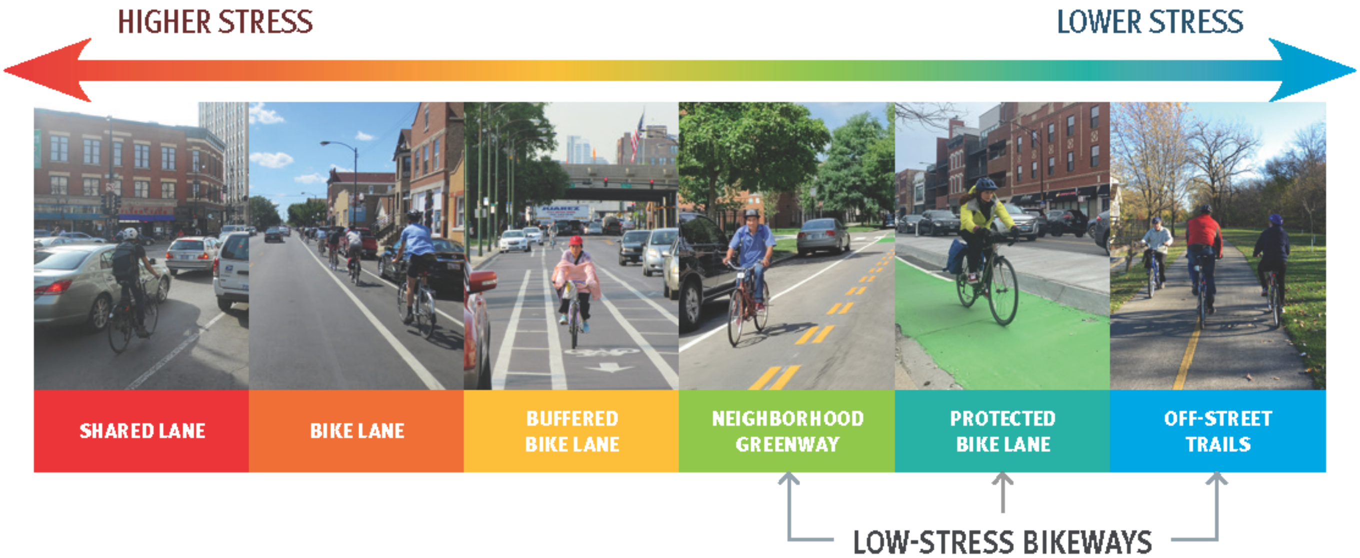

Figure 11: Diagram of different bikeway types and their role in the street grid. Image from CDOT.

In order to create an experiment out of these data, we need bike lanes where we know the year of their installation. In 2021, the Chicago Department of Transportation (CDOT) announced what they described as their "biggest bike lane expansion" with plans to construct nearly 90 miles of bike lane. They issued a map with their press release indicating all of the bike lanes they planned to install in that year. [Cit21b][Cit21a]. This plan focused heavily on constructing bike lanes in underserved sections of the city, such as the West Side neighborhoods of Belmont Cragin, Austin, and North Lawndale, as well as the Far South Side (a diagram can be seen in Figure 13). As these neighborhoods are primarily residential, it focused largely on neighborhood greenways. Of these 90 miles of bikeways, roughly 70 were built, and a full list of the bikeways that we considered for our study are in Table 3.

Categorization	Start Address	End Address	Distance
Bike lane	6100 N Jersey Ave	5600 N Jersey Ave	0.63
Bike lane	5600 N Kedzie Ave	3600 N Kedzie Ave	0.63
Protected bike lane	5730 N Clark St	6400 N Clark St	0.84
Bike lane	6000 W Lawrence Ave	5600 W Lawrence Ave	0.50
Bike lane	3600 W Lawrence Ave	2950 W Lawrence Ave	0.82
Greenway	4700 N Kenmore Ave	5000 N Kenmore Ave	0.38
Greenway	4700 N Winthrop Ave	5000 N Winthrop Ave	0.38
Not on map	2740 W Leland Ave	1500 W Leland Ave	1.55
Bike lane	3400 W Montrose Ave	1430 W Montrose Ave	2.46
Protected bike lane	4000 N Clark St	4400 N Clark St	0.50
Marked shared lanes	4000 N Broadway	4400 N Broadway	0.50
Bike lane	2440 N Sacramento Ave	3400 N Sacramento Ave	1.20
Buffered bike lane	3400 W Diversey Ave	2800 W Diversey Ave	0.75
Not on map	2624 N Milwaukee Ave	3940 N Milwaukee Ave	0.40
Bike lane	2940 N Milwaukee Ave	3200 N Milwaukee Ave	0.33
Greenway	6400 W Roscoe St	4132 W Roscoe St	2.84
Not on map	6400 W Belmont Ave	3840 W Belmont Ave	3.20
Bike lane	6400 W Diversey Ave	5200 W Diversey Ave	1.50
Greenway	6400 W Wrightwood Ave	4000 W Wrightwood Ave	3.00
Buffered bike lane	6400 W Grand Ave	5000 W Grand Ave	1.75
Greenway	5000 W Armitage Ave	2500 W Armitage Ave	3.13
Greenway	1230 W Dickens Ave	320 W Dickens Ave	1.14
Marked shared lanes	2200 N Austin Ave	3600 N Austin Ave	1.75
Bike lane	2400 N Central Ave	3600 N Central Ave	1.50
Bike lane	1500 N Laramie Ave	3600 N Laramie Ave	2.63

Categorization	Start Address	End Address	Distance
Not on map	600 S Laramie Ave	1500 N Laramie Ave	2.38
Greenway	2600 N Kilbourn Ave	3600 N Kilbourn Ave	1.25
Not on map	6400 W Bloomingdale Ave	5200 W Bloomingdale Ave	1.50
Greenway	6000 W Le Moyne St	5000 W Le Moyne St	1.25
Greenway	6000 W Hirsch St	5000 W Hirsch St	1.25
Not on map	400 N Menard Ave	1800 N Menard Ave	2.00
Greenway	300 S Menard Ave	400 N Menard Ave	0.75
Greenway	600 S Lavergne Ave	1400 N Lavergne Ave	2.25
Not on map	4800 W Division St	3600 W Division St	1.50
Marked shared lanes	6000 W Jackson Blvd	5200 W Jackson Blvd	1.00
Bike lane	5600 W Harrison St	4400 W Harrison St	1.50
Greenway	2200 S Keeler Ave	300 N Keeler Ave	2.38
Protected bike lane	1344 S Independence Blvd	100 N Hamlin Blvd	1.29
Greenway	2200 S Independence Ave	1344 S Independence Blvd	0.84
Greenway	1900 S Ridgeway Ave	1400 S Ridgeway Ave	0.56
Greenway	3731 W 19th St	3800 W 19th St	0.09
Greenway	4400 W 14th St	3800 W 14th St	0.75
Greenway	4000 W Polk St	2800 W Polk St	1.25
Protected bike lane	3800 W Douglas Blvd	3100 W Douglas Blvd	0.75
Bike lane	4500 W 16th St	3100 W 16th St	1.75
Bike lane	2200 S Homan Ave	300 S Homan Ave	1.75
Bike lane	1000 N Kedzie Ave	1600 N Kedzie Ave	0.75
Not on map	330 N Kedzie Ave	1000 N Kedzie Ave	0.84
Bike lane	300 S Kedzie Ave	330 N Kedzie Ave	0.66
Not on map	2200 S Kedzie Ave	300 S Kedzie Ave	1.75
Buffered bike lane	3200 W Cermak Rd	2800 W Cermak Rd	0.50
Bike lane	4000 W 26th St	3000 W 26th St	1.25
Protected bike lane	3200 W Lake St	2000 W Lake St	1.50
Protected bike lane	2400 N Damen Ave	2800 N Damen Ave	0.50
Protected bike lane	3200 N Campbell Ave	2400 W Roscoe Ave	0.38
Marked shared lanes	3000 S Loomis St	2700 S Loomis St	0.38
Protected bike lane	300 N Halsted St	500 N Halsted St	0.25
Bike lane	200 W Kinzie St	0 E/W Kinzie St	0.25
Not on map	1800 S Wentworth Ave	1519 S Wells St	0.31
Protected bike lane	1200 S Wells St	900 S Wells St	0.25
Not on map	900 S Wells St	600 S Wells St	0.25
Protected bike lane	1000 W Taylor St	500 W Taylor St	0.63
Protected bike lane	1932 W Harrison St	1800 W Harrison St	0.17
Bike lane	5500 S Racine Ave	5200 S Racine Ave	0.38
Protected bike lane	400 E Garfield Blvd	800 E 55th St	0.50
Bike lane	100 W 71st St	0 E/W 71st St	0.13
Not on map	800 E 71st St	2000 E 71st St	2.00
Buffered bike lane	2000 E 71st St	2400 E 71st St	0.50
Bike lane	7600 S King Drive	6700 S King Drive	1.13
Greenway	1400 W 81st St	1200 W 81st St	0.25
Greenway	1600 W 82nd St	1400 W 82nd St	0.25
Marked shared lanes	4800 W 83rd St	4000 W 83rd St	1.00
Bike lane	4000 W 83rd St	3624 W 83rd St	0.47
Protected bike lane	0 E/W 83rd St	100 E 83rd St	0.12
Marked shared lanes	11900 S Longwood Dr	9500 S Longwood Dr	3.00
Greenway	9700 S Charles St	9500 S Charles St	0.25
Greenway	1700 W 97th St	1623 W 97th St	0.10
Greenway	10500 S Prospect Ave	9700 S Prospect Ave	1.00
Greenway	1600 W 104th St	1400 W 104th St	0.25
Greenway	10500 S Charles St	10400 S Charles St	0.13
Greenway	1600 W 105th St	1400 W 105th St	0.25
Protected bike lane	1600 W 119th St	800 W 119th St	1.00

Categorization	Start Address	End Address	Distance
Bike lane	800 W 119th St	0 E/W 119th St	1.00
Buffered bike lane	9900 S State St	9500 S State St	0.50
Buffered bike lane	9500 S Cottage Grove Ave	9300 S Cottage Grove Ave	0.20
Protected bike lane	11100 S Doty Ave	10100 S Woodlawn Ave	1.60
Buffered bike lane	10000 S Commercial Ave	9400 S Commercial Ave	0.75
Buffered bike lane	11400 S Ewing Ave	10600 S Ewing Ave	1.00
Protected bike lane	11800 S Ewing Ave	11400 S Ewing Ave	0.50
Buffered bike lane	3400 E 112th St	4100 E 112th St	0.88

Table 3: List of bike lanes planned to be installed in 2021, from [Cit21b]. In total, the city installed the following bikeways (based on their classification in [Nei24]: Marked shared lanes: 7.63 miles, Bike lanes: 21.72 miles, Buffered bike lanes: 6.83 miles, Neighborhood greenways: 25.97 miles, Protected bike lanes: 9.94 miles, Not on map: 17.68 miles. Distances estimated based on Chicago street grid layout. Note that some of these bike lanes may have been upgraded to a better classification since 2021, as the cycling map [Nei24] was accessed in December 2024.

We subdivided the total crash dataset by using the STREET_NO, STREET_NAME and STREET_DIRECTION variables based on all of the streets in each bikeway. Then, we combined all of the subsets corresponding to one bikeway category into a single dataframe, such as bikelane.csv or bufferedbikelane.csv. Based on these sub data-frames, we calculated the total number of crashes, the number of injury/tow away crashes, the number of crashes with cyclists, and the number of Dooring crashes by year. Doorings are a crash category where a driver opens their car door into the path of a cyclist, and are considered to be extremely dangerous for cyclists, as the drivers often do not pay attention when opening their door, and the door does not break when hit by a cyclist's momentum, giving the cyclist essentially the experience of running into a stationary wall.[Chi22a].

We can see the results of this analysis in Figure 12. One thing that becomes immediately apparent with the analysis is that cyclist-involved crashes do not seem to have decreased with the installation of bike lanes, but in fact they may have increased, with 86.5 cyclist-involved crashes and 14 doorings per year across 2018-2021, and 113.3 crashes and 12.7 doorings from 2022-2024. The results for the subcategories are often too small to get data that is too significant. On the surface, this would seem to indicate that bike lanes do not actually improve cyclist safety across the city.

However, there are certainly a number of confounding variables that affect the findings which cannot be observed from crash totals alone. Firstly, these results need contextualizing based on the total number of bike trips. As Chicago constructs more bike infrastructure and due to economic pressures and increased emphasis on physical fitness, the number of cycling trips in Chicago increased by 119% from 2019 to 2023 [Rep] (which can be seen in Figure 13). If the total number of cycling trips in Chicago more than doubled from 2019 to 2023, the citywide total cyclist involved crashes increasing only from 1884 to 1936 is some testament that infrastructure improvements are helping. While this includes more than just our study area, the main network improvements in 2021, Figure 13 shows us that Belmont Cragin cycling is up 80-115%, Austin 40-80%, North Lawndale 115-130% and many of the Far South Side neighborhoods 130-170%, which were the main neighborhoods in the study area. Furthermore, the fact that the city publishes a list of bikeways and a travel-planning tool using the existing bikeway system[Nei24] tend to funnel cyclists more onto those routes, so the percentage increase in cyclists along these routes is probably higher than on a randomly selected street in the community.

On the other hand, if publishing this map is funneling cyclists onto potentially dangerous mixed-traffic bikeways, then this bike map may contribute to cycling danger. Another potential problem is lack of enforcement on marked bikeways. Articles like [Chi22a] describe the Milwaukee Avenue bikeway, which is supposed to be a portected bike lane along much of its distance, as being full of parked cars, delivery trucks, and debris, along with many of the protective bollards being knocked over. The city planned to upgrade all bike lanes they classified as protected with concrete curbs that are more durable, but these plans were substantially delayed, with the city citing a concrete worker's strike [Mye22]. Furthermore, bike lane protections do not apply to places like intersections, where the cyclist has to cross the intersection in mixed traffic regardless of the type of cycleway.

5 Conclusion

Across all of the crash datasets and models that we considered, the chief factor in crash severity was relative momentum of the vehicles involved in the crash. For instance, in Figure 4, we see variables like First Crash Type: Sideswipe Same Direction, Rear to Rear, and Parked Motor Vehicle as the least significant, while variables like First Crash Type: Pedestrian and Cyclist being the most significant of severe crashes. The same is true for the causal double LASSO in Figure 7a .Variables including posted speed limit were also highly significant in Figure 5, along with number of

Variable	16	17	18	19	20	21	22	23	24
Total Crashes	95	205	277	281	224	216	243	237	248
Injury or Tow	6	56	79	86	70	76	89	78	80
Cyclist Involved	1	4	9	7	2	5	6	6	5
Doorings	0	0	0	1	0	1	1	3	1
Total Crashes	699	1315	1925	1863	1570	1805	1836	1901	1667
Injury or Tow	96	306	587	542	556	646	599	593	541
Cyclist Involved	14	42	44	39	36	36	46	64	63
Doorings	5	10	8	7	8	7	5	6	12
Total Crashes	204	370	571	559	423	498	505	510	430
Injury or Tow	26	77	155	136	123	131	146	144	124
Cyclist Involved	3	4	8	11	9	6	8	8	13
Doorings	2	0	1	0	0	0	0	0	1
Total Crashes	195	464	684	693	639	731	614	652	569
Injury or Tow	17	112	206	198	190	231	190	200	169
Cyclist Involved	4	12	26	23	8	21	8	17	29
Doorings	1	1	8	5	2	0	0	2	2
Total Crashes	323	572	788	825	712	755	787	778	732
Injury or Tow	27	137	286	276	284	291	278	290	261
Cyclist Involved	2	8	15	15	11	15	18	19	30
Doorings	2	0	2	2	1	3	0	1	4
Total Crashes	1516	2926	4245	4221	3568	4005	3985	4078	3646
Injury or Tow	172	688	1313	1238	1223	1375	1302	1305	1175
Cyclist Involved	24	70	102	95	66	83	86	114	140
Doorings	10	11	19	15	11	11	6	12	20
Cyclist Involved	373	1286	1916	1884	1367	1537	1686	1936	2079
Doorings	75	181	337	259	188	198	184	226	238

Figure 12: Bike lanes installation analysis by year. Red: Total crashes, crashes resulting in injury or tow, crashes involving cyclists, and doorings for 2021 marked shared lanes installations. Orange: Bike lane installations. Yellow: Buffered bike lane installations. Lime: Neighborhood greenway installations. Green: Protected bike lane installations. Blue: Totals for all 2021 bike lane installations. Pink: Citywide cyclist involved crashes and doorings. Note this can include crashes on designated bikeways as well as others.

vehicles involved in the crash. Our recommendation to the city would be finding more ways to separate motorists and pedestrians/cyclists, as well as potentially reducing the speed limits on residential roads.

Regarding the traffic improvements in Section 4, we observe mixed results. Although the pedestrian crashes in the LTTC zones were decreased, some confounding variables make concluding anything too significant, especially in the River North study area, difficult. The crashes dataset was also not sufficient to test some of the headline conclusions regarding LTTC, such as percentage of drivers yielding. The bike lanes saw an increase in bike crashes compared to pre-COVID, although this was confounded by a doubling in total bike trips. When this is factored in, it does seem like separating cyclists and motorists improves both of their safety.

References

[Chi] Chicago Police Department. *Police Area, District and Beat Boundaries*. URL: https://chicagopd.hub.arcgis.com/documents/ChicagoPD::area-district-beat-11x17-1/explore.

[Chi19] RNRA Chicago. *New 42nd Ward Traffic Safety Infrastructure: Left Turn Traffic Calming along State Street, High Crash Corridor*. 2019. URL: https://rnrachicago.org/new-42nd-ward-traffic-safety-infrastructure-left-turn-traffic-calming-along-state-street-high-crash-corridor/.

[Chi22a] Block Club Chicago. *Life and Death in Chicago's Most Dangerous Bike Lane*. 2022. URL: https://blockclubchicago.org/2022/11/29/life-and-death-in-chicagos-most-dangerous-bike-lane/.

[Chi22b] Streetsblog Chicago. *Eyes on the Street: CDOT Hardens Left Turns on Ashland with Speed Bumps and Bollards*. 2022. URL: https://chi.streetsblog.org/2022/09/16/eyes-on-the-street-cdot-hardens-left-turns-on-ashland-with-speed-bumps-and-bollards.

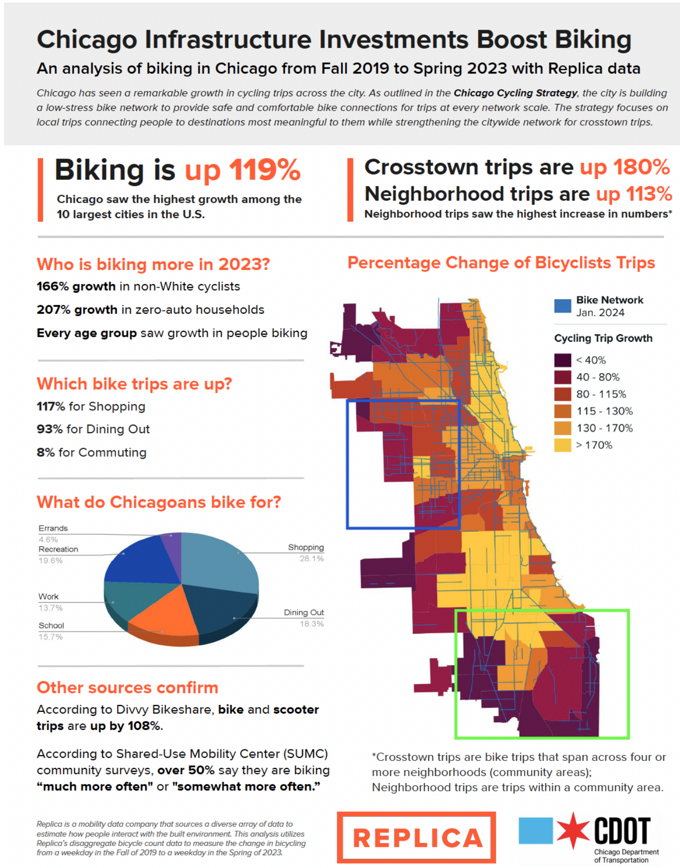

Figure 13: Chicago biking analysis report conducted by Replica for the Chicago Department of Transportation. It shows proportion change in cycling trips by community area from 2019 to 2023. The areas that were focuses in the 2021 bike lane construction are in the boxes. Blue: West Side: Belmont-Cragin, Austin, North Lawndale, Green: Far South Side. [Rep]

[Chi23] City of Chicago Department of Transportation. *Left Turn Traffic Calming Plan*. 2023. URL: https://www.chicago.gov/content/dam/city/sites/complete-streets/pdfs/left-turn-traffic-calming-01232023.pdf.

[Cit] City of Chicago Data Portal. *Traffic Crashes Data*. URL: https://data.cityofchicago.org/Transportation/Traffic-Crashes-Crashes/85ca-t3if/about_data.

[Cit21a] City of Chicago. *CDOT Announces Biggest Bike Lane Expansion in City's History*. 2021. URL: https://www.chicago.gov/city/en/depts/cdot/provdrs/bike/news/2021/september/cdot-announces-biggest-bike-lane-expansion-in-city-s-history-thr.html.

[Cit21b] City of Chicago. *Chicago Community Cycling Plan*. 2021. URL: https://www.chicago.gov/content/dam/city/depts/cdot/bike/2021/Chicago%20Community%20Cycling_2021-09-21.pdf.

[Mye22] Myers, Quinn. *Quarry Workers Strike Stalls Milwaukee Avenue Resurfacing Work — And Other Street, Sidewalk Plans Across The City*. 2022. URL: https://blockclubchicago.org/2022/07/11/milwaukee-avenue-resurfacing-project-quarry-workers-strike/.

[Nei24] Center for Neighborhood Technology. *Chicago Bike Routes*. 2024. URL: https://apps.cnt.org/bikechi/.

[Rep] Replica. *Chicago Infrastructure Investments Boost Biking*. URL: https://www.replicahq.com/post/measuring-chicagos-boost-in-biking.