Exploring Chicago Traffic Crashes Data in Tableau

CIS 5270 Professor Shilpa Balan

Yangyang Jia (yjia12@calstatela.edu)

Department of Information Systems, California State University Los Angeles

TABLE of CONTENTS

1. **Dataset URL**

2. **Data Cleaning**

3. **Data Visualizations & Explanations for the Analysis questions.**

4. **Dashboard**

5. **Story Telling**

A) Data set URL's:

Traffic Crashes – Vehicles:

https://data.cityofchicago.org/Transportation/Traffic-Crashes-Vehicles/68nd-jvt3

This dataset contains information about vehicles involved in a traffic crash from 2015 to 2018. "Vehicle" information includes motor vehicle and non-motor vehicle modes of transportation, such as bicycles and pedestrians.

Traffic Crashes – People:

https://data.cityofchicago.org/dataset/Traffic-Crashes-People-Dashboard/7fud-yfx4

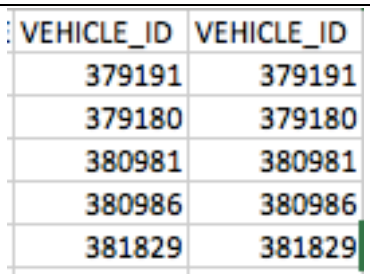
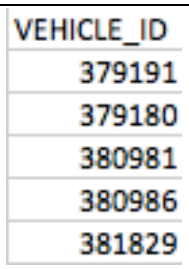
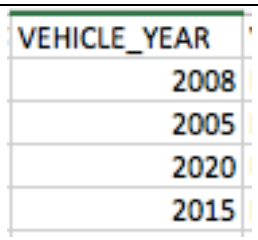
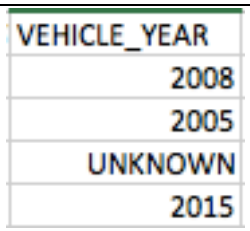
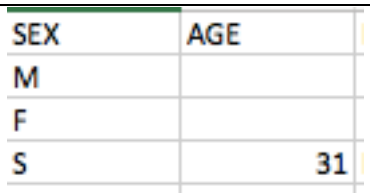
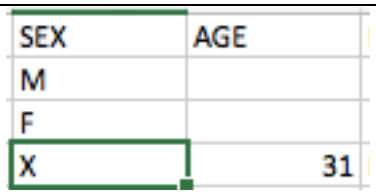
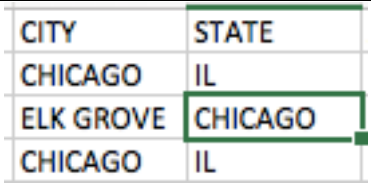
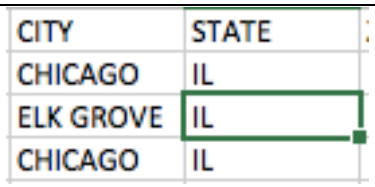
This dataset contains information about people involved in a crash and if any injuries were sustained from 2015 to 2018. Some people involved in a crash may not have been an occupant in a motor vehicle, but may have been a pedestrian, bicyclist, or using another non-motor vehicle mode of transportation.

Traffic Crashes – Crashes:

https://data.cityofchicago.org/Transportation/Traffic-Crashes-Crashes-Dashboard/8tdq-a5dp

This dataset shows information about each traffic crash on city streets within the City of Chicago limits and under the jurisdiction of Chicago Police Department (CPD) from 2015 to 2018. Many of the crash parameters, including street condition data, weather condition, and posted speed limits, are recorded by the reporting officer based on best available information at the time.

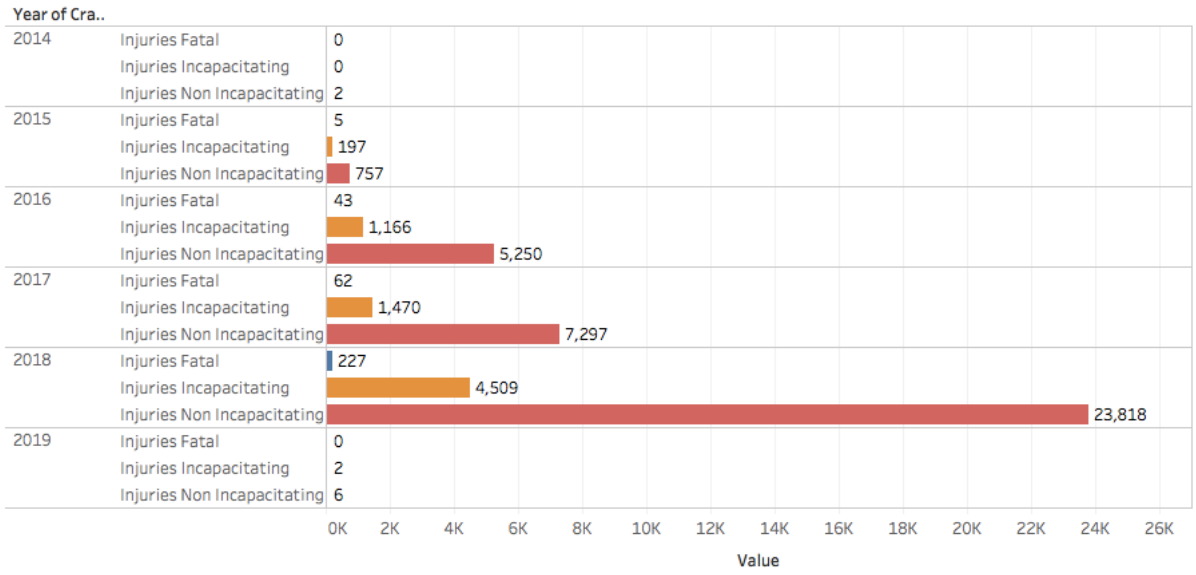
B) Data Cleaning:

Score / Problem	Dirty Data	Cleaned Data / Remarks
1. Duplicated Record	 Vehicle_ID is duplicated.	 Delete one of the duplicated columns.
2. Contradicting Records	 2020 is contradicting the records.	 Correct the record to be unknown
3. Illegal values	 “S” doesn’t represent anything for SEX	 Changed “S” to “X” as unknown value.
4. Misfielded values	 “Chicago” as a city name shouldn’t show on the STATE filed.	 Corrected “Chicago” to “IL” as Chicago is in IL State

5. Embedded values	VEHICLE_ID 10 08/04/2015 12:40:00 PM 96 07/31/2015 05:50:00 PM 954 09/02/2015 11:45:00 AM It combined Vehicle_ID and Crashes Time to one Column	VEHICLE_ID | CRASH_DATE 10 | 08/04/2015 12:40:00 PM 96 | 07/31/2015 05:50:00 PM 954 | 09/02/2015 11:45:00 AM 9561 | 10/31/2015 09:30:00 PM Separated two values to two different columns.

C) Data Visualizations:

I What is the count of total number of fatal, incapacitating and non-incapacitating injuries in each year from 2014 - 2019?



[Tools used: Dates (year), Injuries Fatal, Injuries Incapacitating, Injuries Non-Incapacitating, Lines]

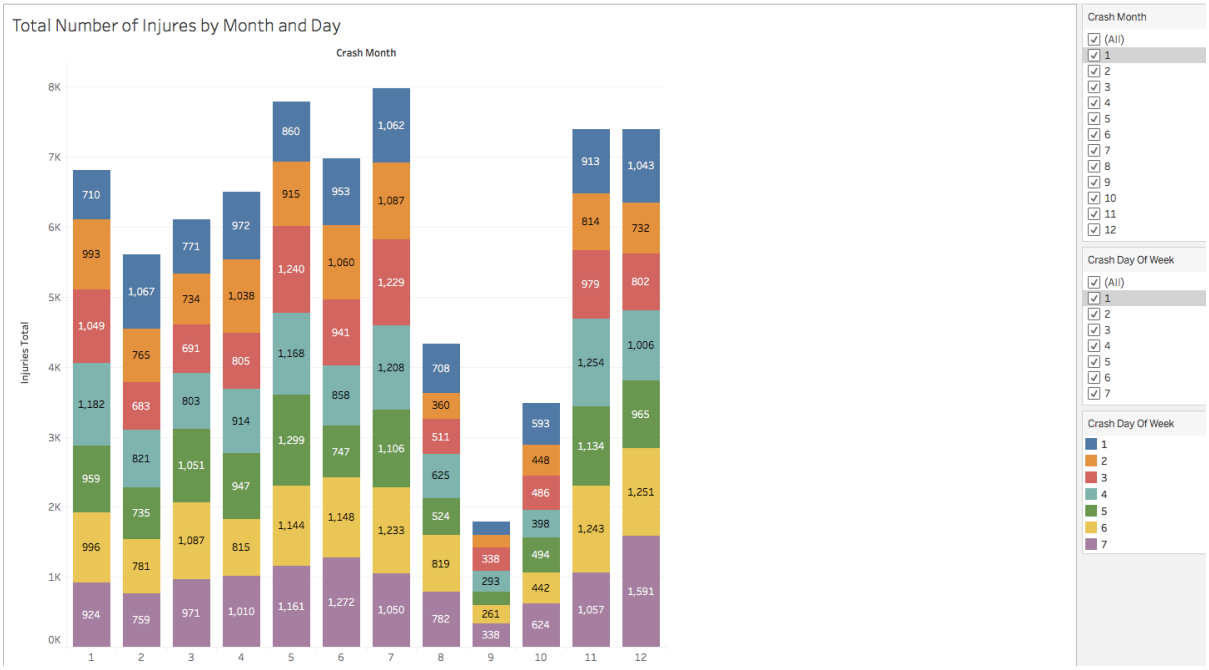
The above chart table shows the following measures for each year form 2014 – 2019.

- Total Number of Fatal Injuries.
- Total Number of Incapacitating Injuries.
- Total Number of Non-Incapacitating Injuries.

This helps us know the count of the total number of accidents based on the different level of injuries occurred per year from 2014 – 2019. From the chart, we can tell that 2018 reached to

the highest records of total number of accidents. This leads us for a deeply research and found out what cause 2018 reached the highest total numbers.

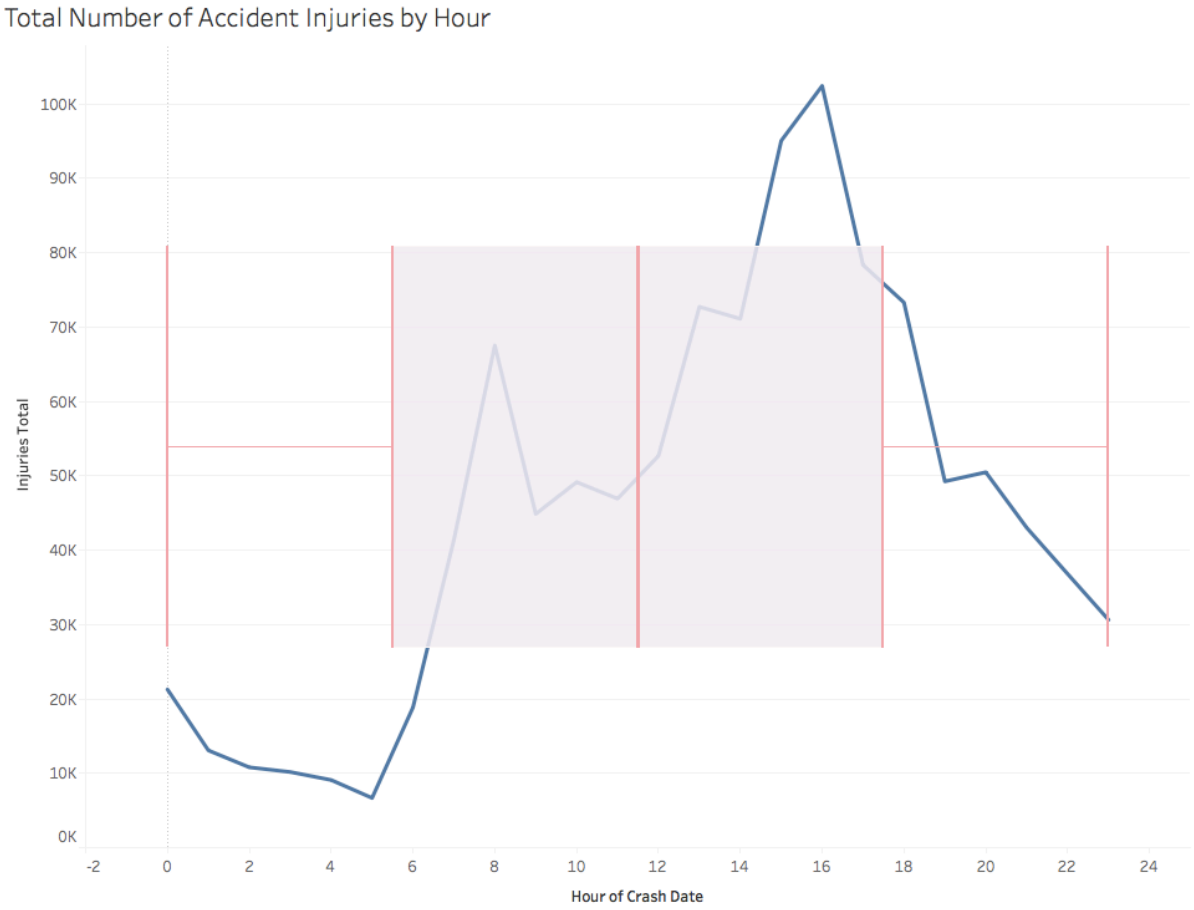
II What is the count of total number of injuries by weekday from 2014 - 2019?



[Tools used: Dual Axis Chart, Stacked Bars]

The above analysis using a stacked bar chart shows the count of total number of injures accident that have happened by monthly and daily from 2014 – 2019. The different day are all displayed in different colors as shown in the filter label. This helps us understand that July, May and December are the top three month of highest total number of injuries from 2014 – 2019. Besides, Sunday, Wednesday and Tuesday are the top three day of highest total number of injuries from 2014 – 2019.

III What is the count of total number of injuries by hourly from 2014 - 2019?



[Tools used: Lines (Continuous), Reference Line, Box Plot]

The above analysis using a Line chart shows the count of total number of injures accident that have happened by hour from 2014 – 2019. By adding the reference line (Box Plot) shows the data within 1.5 times the IQR. So, this helps us easily understand that from 6am to 18pm are the peak time that accident occurred. Especially, 16pm reached to the highest count of total number of accident injures.

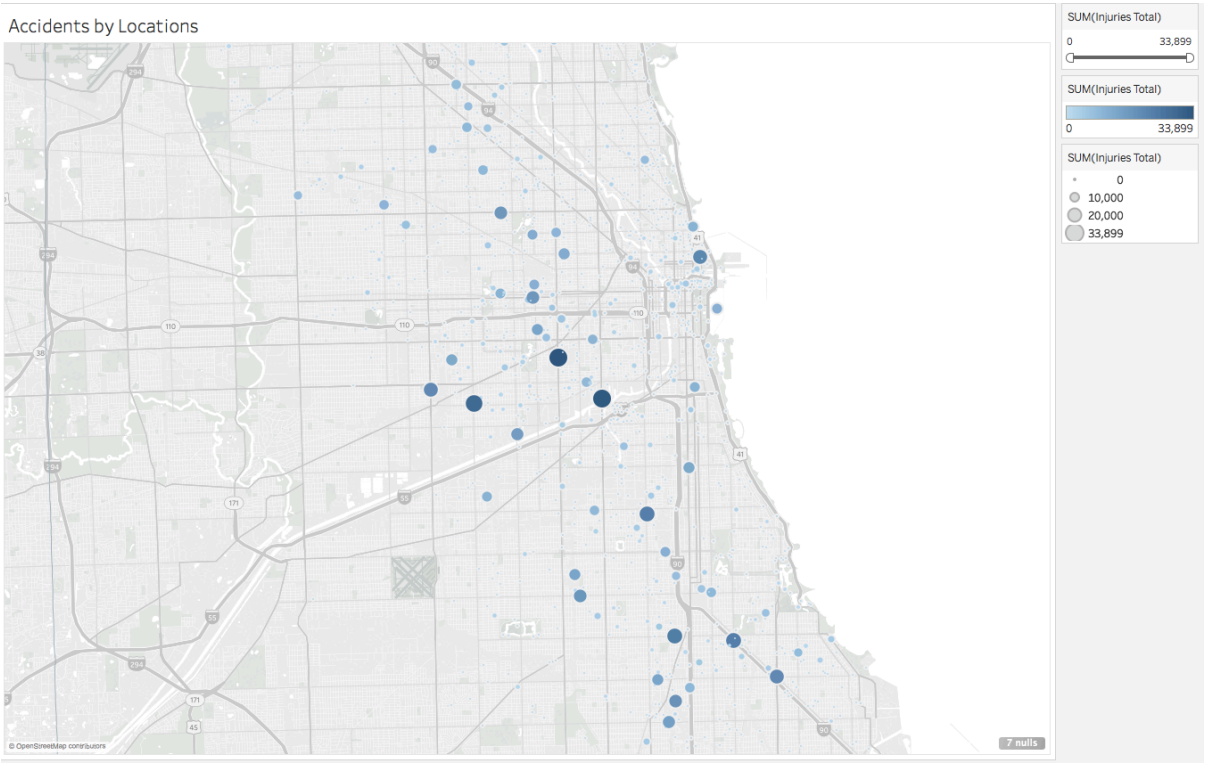
IV What are the contributing factors lead to the accident injuries?

		Roadway Surface Cond						
Lighting Co..	Device Condition	DRY	ICE	OTHER	SAND, M..	SNOW O..	UNKNO..	WET
DARKNESS	FUNCTIONING IMPROPERLY	106	0			0	6	48
	FUNCTIONING PROPERLY	2,221	84	4	0	210	111	638
	NO CONTROLS	1,694	13	2	5	192	67	724
	NOT FUNCTIONING	25					0	0
	OTHER	13		4		0	0	8
	UNKNOWN	185	0	0		34	91	69
	WORN REFLECTIVE MATERIAL	0						
DARKNESS, LIGHTED ROAD	FUNCTIONING IMPROPERLY	251	0			66	0	71
	FUNCTIONING PROPERLY	20,032	46	14	4	1,545	311	5,815
	MISSING	6						
	NO CONTROLS	11,071	224	42	0	580	362	2,443
	NOT FUNCTIONING	99	0		0	4	6	141
	OTHER	251	1	12	1	64	0	127
	UNKNOWN	1,351	1	6	6	72	504	408
	WORN REFLECTIVE MATERIAL	20	0				0	0
DAWN	FUNCTIONING IMPROPERLY	4				0		22
	FUNCTIONING PROPERLY	1,025	39	8		8	12	384
	NO CONTROLS	672	222	4	0	35	22	231
	NOT FUNCTIONING	0	0			0	0	0
	OTHER	6		0		0	0	22
	UNKNOWN	150	0			0	13	37
	WORN REFLECTIVE MATERIAL	0						
DAYLIGHT	FUNCTIONING IMPROPERLY	795	9			44	12	260
	FUNCTIONING PROPERLY	43,519	494	131	6	1,217	936	8,658
	MISSING	14						8
	NO CONTROLS	33,785	203	71	12	1,132	617	5,387
	NOT FUNCTIONING	484	0	0	0	6	8	22
	OTHER	853	24	4	0	10	8	83
	UNKNOWN	1,697	4	12	0	77	659	456
	WORN REFLECTIVE MATERIAL	77				6	0	0
DUSK	FUNCTIONING IMPROPERLY	12				0		0
	FUNCTIONING PROPERLY	1,993	4	12	0	83	84	292
	NO CONTROLS	2,399	86	4		101	30	360
	NOT FUNCTIONING	6				0		2
	OTHER	32	0	0			4	9
	UNKNOWN	189	0	0		0	16	40
	WORN REFLECTIVE MATERIAL	0						0

[Tools used: Highlight Table, Group]

The above analysis using a highlight table chart shows that the different contributing factors which cause the accident injures from 2014 – 2019. The darker blue color shows the greater count of total number of injuries. This helps us to know that daylight of Lighting condition, functioning properly and no controls of device condition, dry and wet of roadway surface condition could cause higher records of accident injuries.

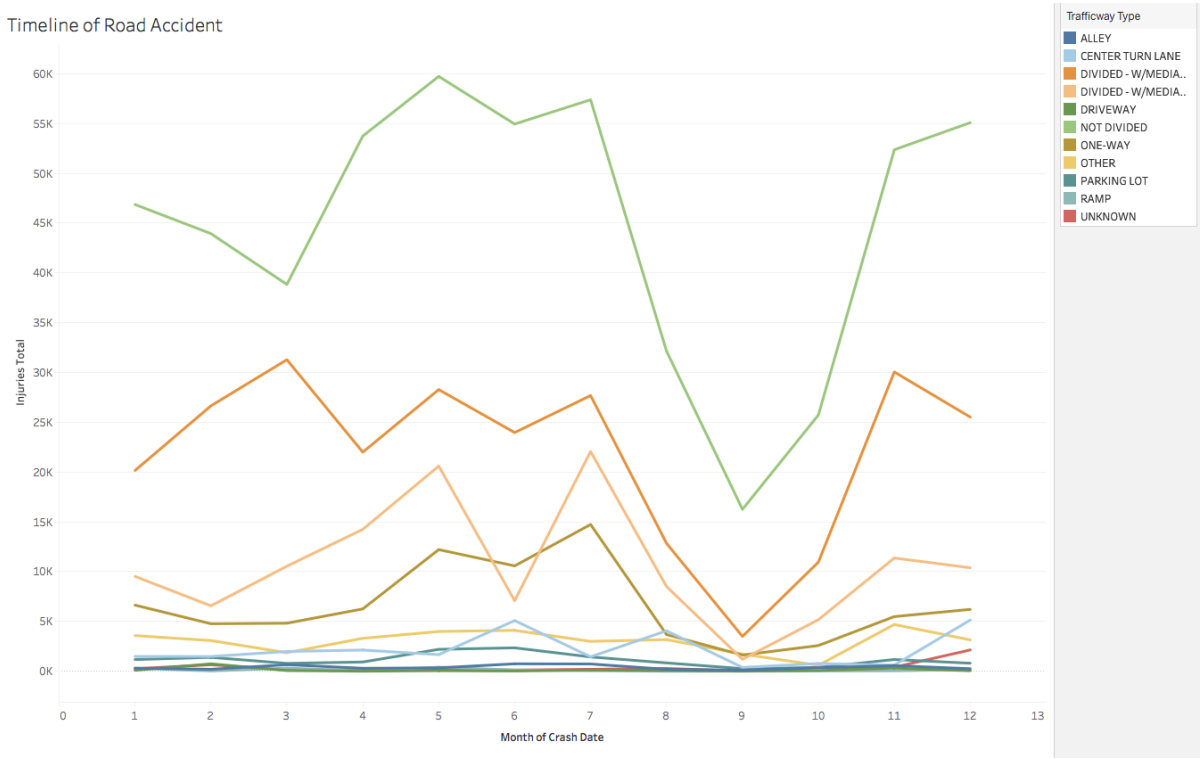
V What are the locations that have the most accident injures occurred from year 2014 – 2019?



[Tools used: Geographic Maps]

The above analysis using a symbol map table shows that the different locations that have the different count of total number of the accident injures from 2014 – 2019. From the table, we could tell, the darker blue and bigger size of circle has greater number of accident injuries. The location (Latitude: 41.8607, Longitude: -87.686) has the highest count of total number of accident injuries which reaches to 33899. While, Latitude: 41.8467, Longitude: -87.666 and Latitude: 41.8450, Longitude: -87.725 reaches to top 2 and 3 highest count of total number of accident injuries. This map simple shows us that the top 3 highest count of total number of accident injuries happen really close to each other, which are within 5 miles to each location.

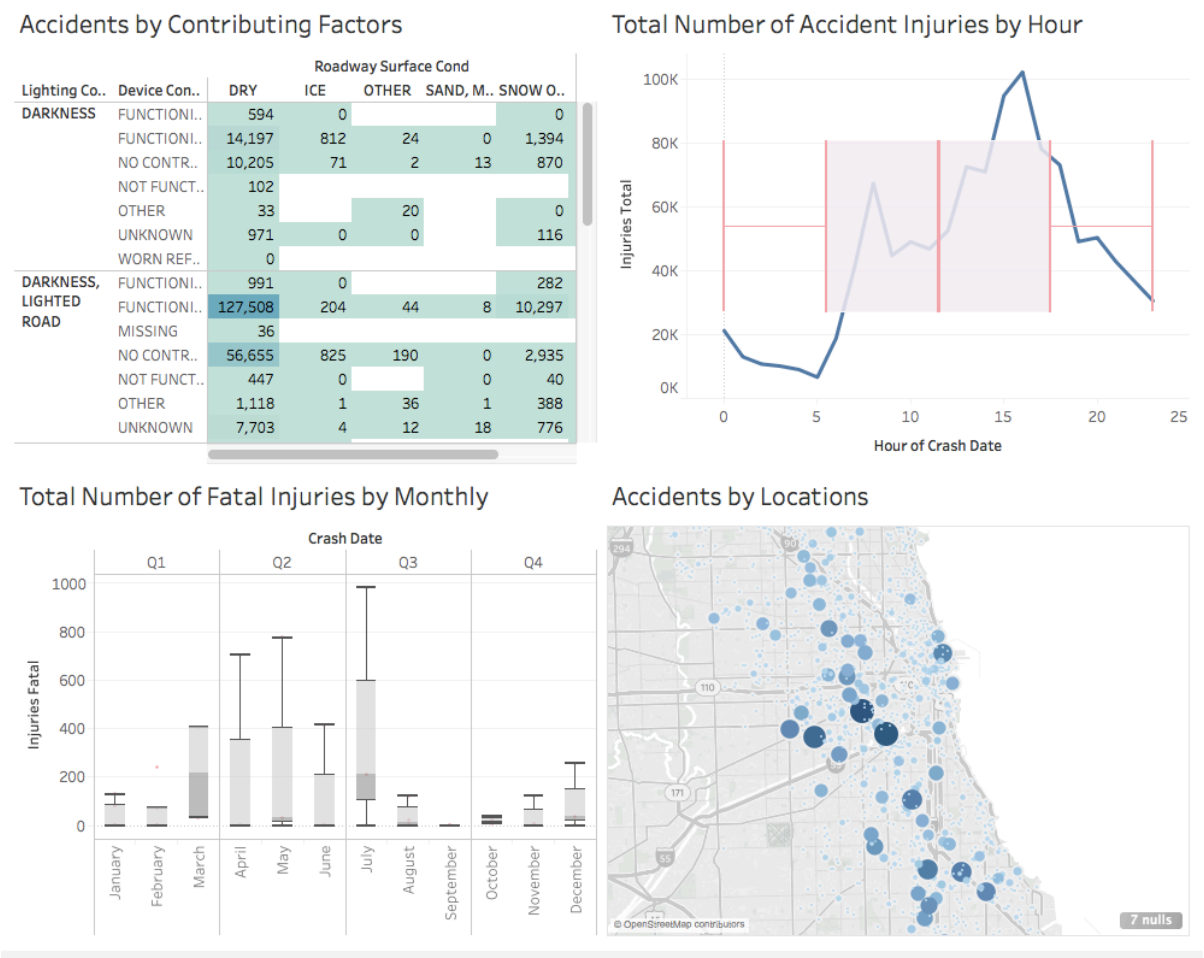
VI What type of road has the higher rate of accident injuries?



[Tools used: Lines (Continuous), Forecast Trend Lines]

The above analysis using a line chart shows that the count of total number of the accident injures that happened in the different types of road by monthly from 2014 – 2019. From this table, we could easily tell which type of road would cause higher number of accident injures. For instance, not divided and Divided –W/Media (not raised) have the top 2 highest count of total number of accident injures. Besides, most of the accidents happened from April to July and November as well.

D) Dashboard



E) Story Telling

The study has considered every aspect of the causative factors leading to traffic accidents, such as the effects of weather, seasonal variation, and road and lighting conditions. The common human errors leading to accidents have been discussed. Other factors, such as the ethnic distribution, and their relation to road accidents have shown the effect of the social structure on the problems. The purpose of this analysis is to explore and gain a better understanding of some of the factors that affect the likelihood of vehicle crashes.

First, Figure 1 demonstrates the view of accident injuries occurrences in a pattern which is corresponding to the common sense. As shown in Fig. 1(a) At the middle and end month of year like May to July, November to December has higher amount of accident occurrences from 2014 -2019. (b) Less accidents happened during the weekend, more accidents happened from Tuesday to Thursday. In summary, the Tuesday, Wednesday and Thursday in May, July, November and December have the highest count of total number accident occurrences that happened among the year.

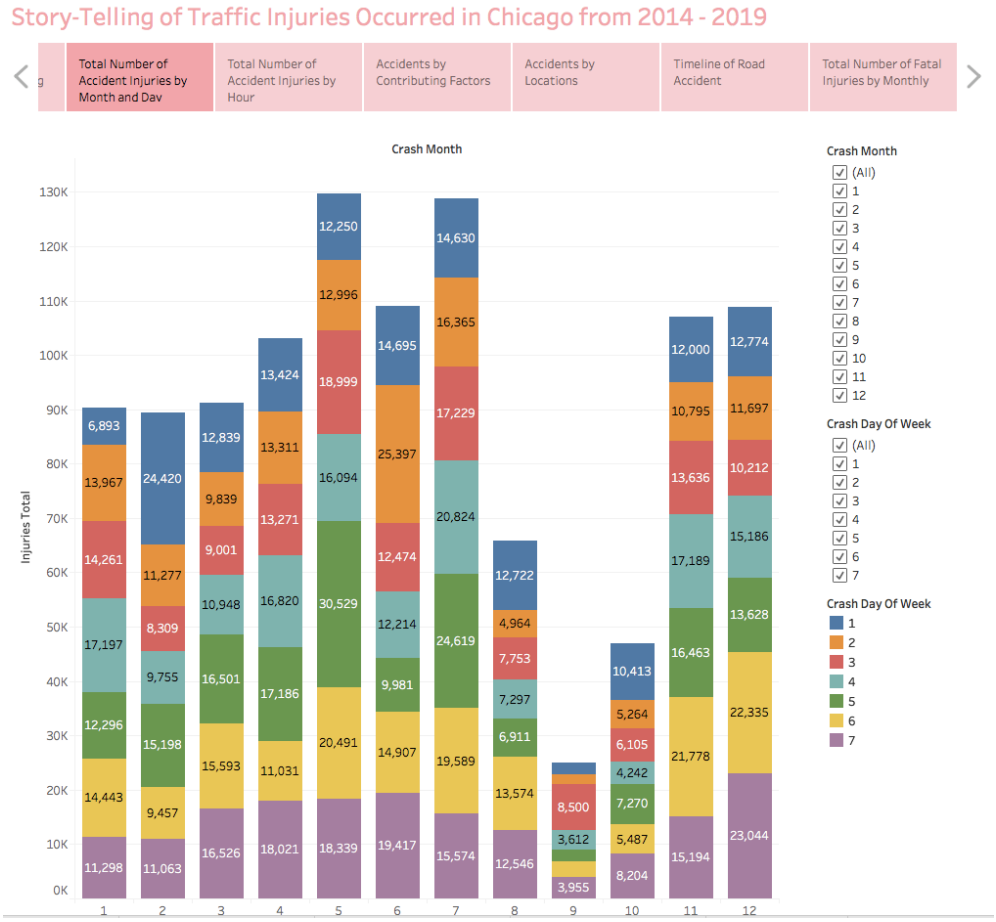


Fig.1. Accident injure occurrences by monthly and daily from 2014 – 2019

Second, I investigated the pattern in accident occurrences by hourly. Figure 2 shows that from 6 am to 6 pm are the peak time that may cause traffic accidents, and have greater total number of accidents injures. For instance, accident injures occurred at 4 pm reaches to the

highest amount which is above 100 thousand from year 2014 to 2019. While, the accident injures occurred at 4 am reaches to the lowest amount which is about 8 thousand. The overtime charts of both periodical patterns, which indicate is that the early peak in the weekend is both smaller and more postponed than those in the weekdays. In summary, the early and evening peak of accidents is closely correlated to the peak hours in roads.

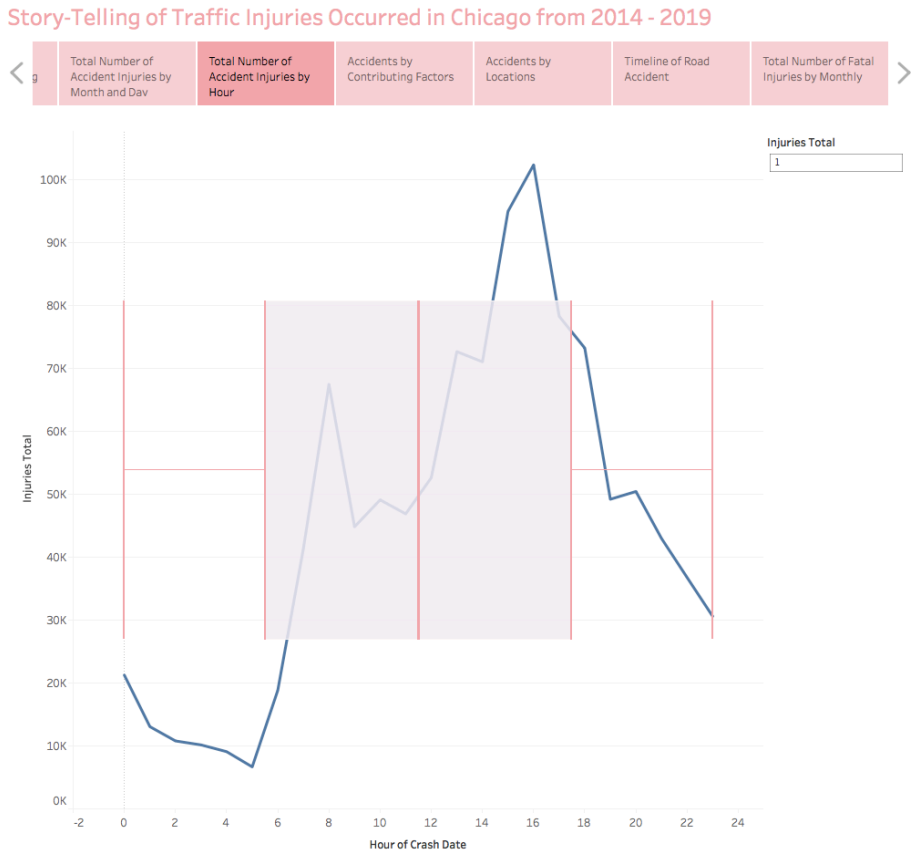


Fig.2. Accident injure occurrences by hourly from 2014 – 2019

Third, I exploit the spatial pattern in accident occurrences with different granularities. Figure 3(a) shows the latitude and longitude coordinates of each accident, and totally 33806 accident occurred in Ashland Ave, Chicago (Latitude: 41.8467, Longitude: -87.666). Figure 3 (b) describes the spatial view of accident occurrences are surrounds in the close areas, Western Ave, Pulaski Rd (the deeper the color, the more the accident in the

region). Figure 3 (c) shows where accidents occurred on the cross zone of two or more freeways.

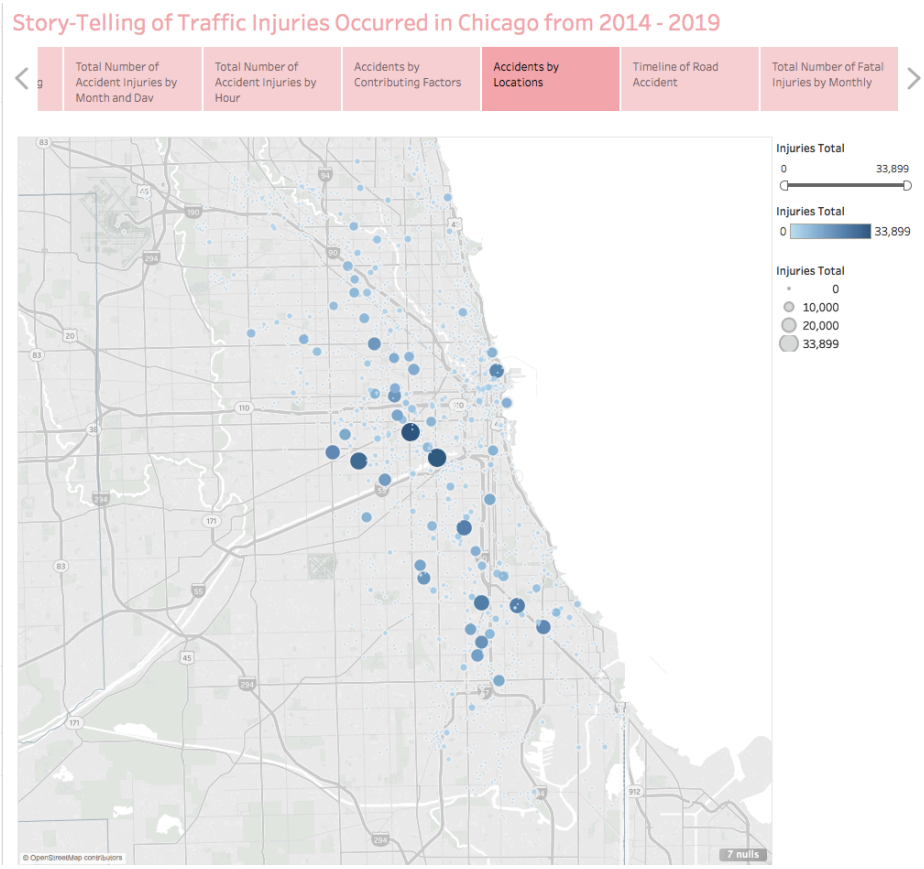
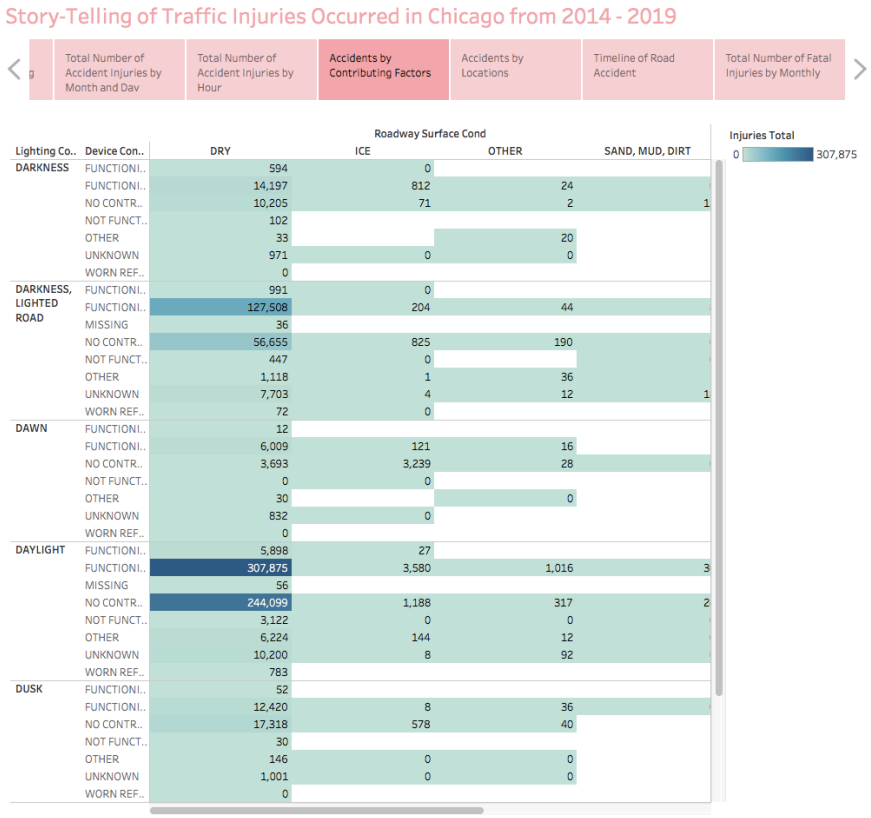


Fig.3. Accident injures by locations

Forth, Figure 4 presents visualization results of crash-type analysis with the feature of lighting condition, device condition and roadway surfaces condition. We first select five types of lighting segments, which would be darkness, darkness-lighted road, dawn, daylight, dusk. Then we have 7 types of device condition, which are functioning improperly, functioning properly, no controls, not functioning, other, unknown and worn reflective material. Last comes to roadway 7 types of surface conditions, dry, ice, other, sand, snow or slush, unknown, wet. We notice that darkness-lighted road and daylight, those two lighting conditions could cause higher number of accident injures, while functioning properly and no controls of device condition have highest accident occurrences. Interesting fact is that, the dry

road cause higher accidents than any other types of roadway surface condition. So, we could know that roadway surface are not the main factor that leads to accident.



Lighting Co..	Device Con..	DRY	ICE	OTHER	SAND, MUD, DIRT
DARKNESS	FUNCTIONI..	594	0		
	FUNCTIONI..	14,197	812	24	
	NO CONTR..	10,205	71	2	1
	NOT FUNCT..	102			
	OTHER	33		20	
	UNKNOWN	971	0	0	
	WORN REF..	0			
DARKNESS, LIGHTED ROAD	FUNCTIONI..	991	0		
	FUNCTIONI..	127,508	204	44	
	MISSING	36			
	NO CONTR..	56,655	825	190	
	NOT FUNCT..	447	0		
	OTHER	1,118	1	36	
	UNKNOWN	7,703	4	12	1
	WORN REF..	72	0		
DAWN	FUNCTIONI..	12			
	FUNCTIONI..	6,009	121	16	
	NO CONTR..	3,693	3,239	28	
	NOT FUNCT..	0	0		
	OTHER	30		0	
	UNKNOWN	832	0		
	WORN REF..	0			
DAYLIGHT	FUNCTIONI..	5,898	27		
	FUNCTIONI..	307,875	3,580	1,016	3
	MISSING	56			
	NO CONTR..	244,099	1,188	317	2
	NOT FUNCT..	3,122	0	0	
	OTHER	6,224	144	12	
	UNKNOWN	10,200	8	92	
	WORN REF..	783			
DUSK	FUNCTIONI..	52			
	FUNCTIONI..	12,420	8	36	
	NO CONTR..	17,318	578	40	
	NOT FUNCT..	30			
	OTHER	146	0	0	
	UNKNOWN	1,001	0	0	
	WORN REF..	0			

Fig.4. Accident injures by crash type (lighting condition, device condition and roadway surfaces condition)

In conclusion. Given the massive traffic accident dataset in Chicago, accident occurrence analysis is known to identify the main factors that contribute to crash type, crash location and severity. By analyzing the traffic accident data from 2014 to 2019 in Chicago, USA, I came up with this research paper of accident occurrence analysis and visualization method in both spatial and temporal dimensions, in order to predict when and where an accident with a certain crash type will happen sequentially by whom. But there is still have room for me to improve with the process of researching deeper and deeper.

References:

1. F.Ahmed Malik, “Road Accidents and Prevention” (2017.04)

2. E.Laiza King, “Top 15 Causes of Car Accidents and How You Can Prevent Them” (2016.08)

3.O.Maureen Donnell, “Exploring NYC Vehicle Crash Data in Tableau” (2015. 08) https://interworks.com/blog/modonnell/2015/08/26/exploring-nyc-vehicle-crash-data-tableau/

4. L.Fan Xiao, “Context-Aware Big Data Analytics and Visualization for City-Wide Traffic Accidents”

5. Jason Chen, “Visualizing NYC Traffic Accidents Before and After Vision Zero” (2017.02)