Final Project by Yuhan Wang

Introduction:

1. What is the dataset?

   The data is "NYS Motor Vehicle Crashes and Insurance Reduction". It is composed of 6 datasets including Facility, Case, Individual, Vehicle, Violation and Insurance, each covering one aspect of a vehicle crash or insurance involved.

   A glance of the dataset is:

   Facility:

	Facility	Year	Total Number of Motor Vehicle Crashes	Number of Fatal Crashes
0	Bayonne Bridge	2000	24	0
1	Bayonne Bridge	2001	19	1
2	Bayonne Bridge	2002	13	0

   Case:

	Year	Crash Descriptor	Time	Date	Day of Week	Police Report	Lighting Conditions	Municipality	Collision Type Descriptor	County Name	Road Descriptor	Weather Conditions	Traffic Control Device	Road Surface Conditions	Ref Lo
0	2014	Property Damage Accident	8:30	2014-01-03T00:00:00	Friday	Y	Daylight	WATERTOWN	OVERTAKING	JEFFERSON	Straight and Level	Clear	None	Snow/Ice	
1	2014	Property Damage & Injury Accident	14:48	2014-01-03T00:00:00	Friday	Y	Daylight	WATERTOWN	REAR END	JEFFERSON	Straight and Level	Cloudy	Traffic Signal	Snow/Ice	730
2	2014	Property Damage & Injury Accident	11:30	2014-01-04T00:00:00	Saturday	Y	Daylight	WATERTOWN	OTHER	JEFFERSON	Straight and Grade	Clear	None	Snow/Ice	

   Individual:

	Year	Case Individual ID	Case Vehicle ID	Victim Status	Role Type	Seating Position	Ejection	License State Code	Sex	Transported By	Safety Equipment	Injury Descriptor	Injury Location	Injury Severity	Age
0	2014	15799027	12117200	Conscious	Driver of a Motor Vehicle in Transport	Driver	Not Ejected	NY	M	Not Entered	Lap Belt/Harness	Complaint of Pain	Shoulder-Upper Arm	Minor	43.0
1	2014	15743949	12076145	Not Applicable	Passenger of a Motor Vehicle in Transport	Passenger - 3 Front Right	Not Ejected	NaN	F	Not Entered	Lap Belt/Harness	Not Applicable	Not Applicable	Uninjured	38.0
2	2014	15444687	11856196	Not Applicable	Driver of a Motor Vehicle in Transport	Driver	Not Ejected	NY	F	Not Entered	Lap Belt/Harness	Not Applicable	Not Applicable	Uninjured	40.0

   Vehicle:

	Year	Case Vehicle ID	Vehicle Body Type	Registration Class	Action Prior to Accident	Type / Axles of Truck or Bus	Direction of Travel	Fuel Type	Vehicle Year	State of Registration	Number of Occupants	Engine Cylinders	Vehicle Make	Contributing Factor 1	Contributing Factor Description
0	2014	12100031	UNKNOWN VEHICLE	Not Entered	Unknown	Not Entered	Unknown	Not Entered	NaN	NaN	NaN	NaN	NaN	HUMAN	Not Entered
1	2014	12638921	TAXI	Not Entered	Slowing or Stopping	Not Entered	South	Not Entered	NaN	NJ	2.0	NaN	NaN	HUMAN	Unknown
2	2014	12011955	UNKNOWN VEHICLE	Not Entered	Unknown	Not Entered	Unknown	Not Entered	NaN	NaN	NaN	NaN	NaN	HUMAN	Not Entered

   Violation:

	Year	Violation Description	Violation Code	Case Individual ID
0	2014	PASSED ON RIGHT	1122A	15943211
1	2014	FOLLOWING TOO CLOSELY	1129A	15551946
2	2014	SPEED NOT REASONABLE & PRUDENT	1180A	16054549

   Insurance:

	Month	Year	Age at Completion	Gender	County of Mailing Address	Sponsor	Delivery Method	Course Completion Count
0	9	2014	58	F	ONEIDA	04 - AARP	CLASSROOM	1
1	1	2015	20	M	BRONX	25 - USA TRAINING COMP	INTERNET	1
2	1	2013	43	F	ULSTER	18 - NY SAFETY PROGRAM	CLASSROOM	1

2. Where did you get it from?
   The data is from Kaggle (https://www.kaggle.com/new-york-state/nys-motor-vehicle-crashes-and-insurance-reduction#motor-vehicle-point-insurance-reduction-program-pirp-participation-five-year-window.csv).

3. Why did you choose this particular data?
   I am interested in what factors might cause higher rate of vehicle crashes, so I'd like to explore this from this data so as to give some suggestions for safer road trip.

4. What types of questions were you hoping to explore with this data?
   As I said before, I'd like to explore the factors influencing vehicle crash probability in order to give some suggestions for safer road trip.

Summary of Data:
Connection Map cannot be plotted from the chosen dataset, so I didn't plot it. The other plots are as follows.
   1. Histogram

Histogram of Ages at the Crash and Age at Completion of Insurance
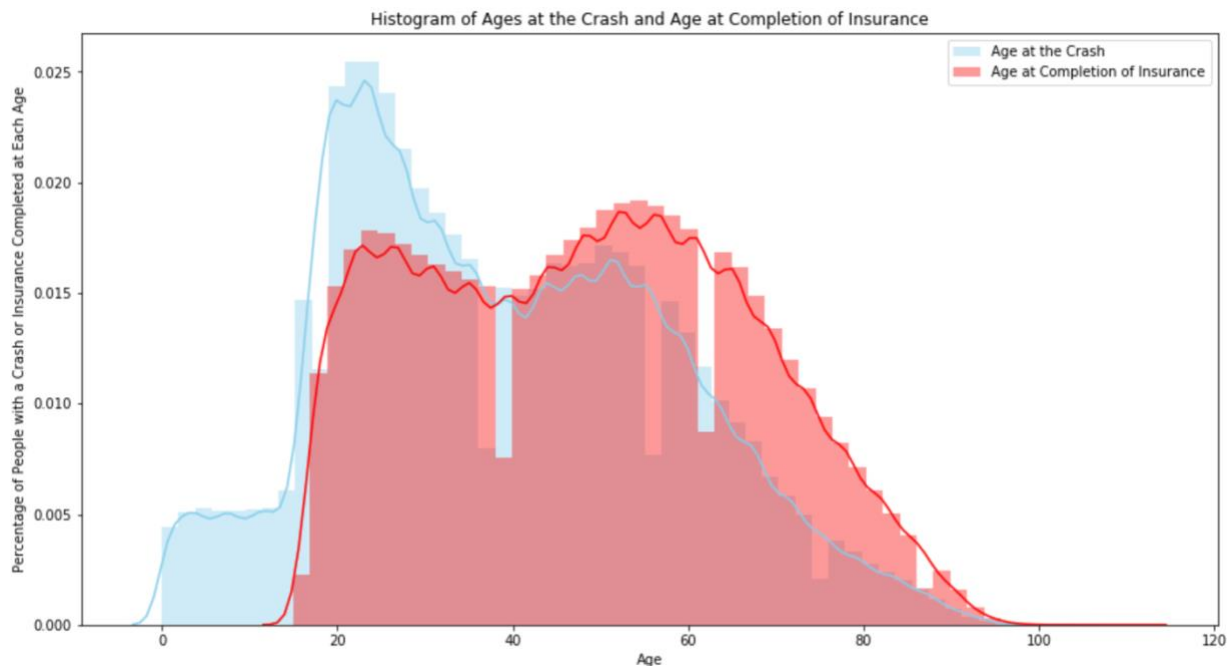
I plotted a histogram of age at the crash and age at completion of insurance. It usually takes years to complete an insurance because there are some lags shown by the two histograms. One thing noticeable is that children are often involved in a crash but with no insurance; another thing is that age around 20s and 50s are two peaks with most vehicle crashes.

   2. Barplot

Probability of Getting Injured Severely when Sitting in Different Positions in Cars, distinguished by Sex
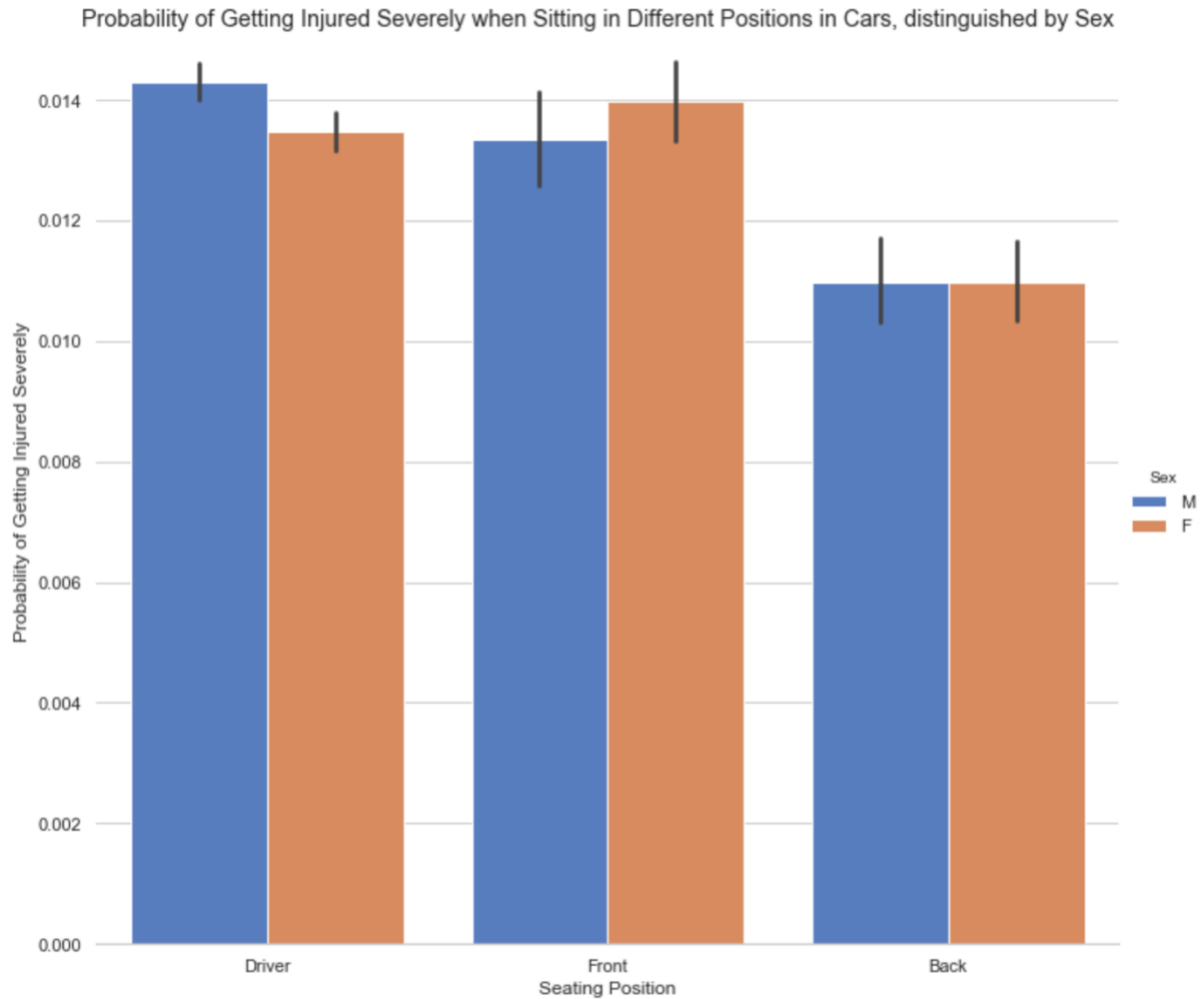

I filtered injury severity to only include 'Severe Injured' and 'Killed', indicating a very severe crash, and plotted a barplot to see whether seating position in a car affects injury severity, distinguished by sex. Results show that driver and front seats are more dangerous but back seats are significantly safer than the others. However, there are no significant difference between genders.

3. Boxplot

Boxplot of Hour at the Crash for each Dat of Week
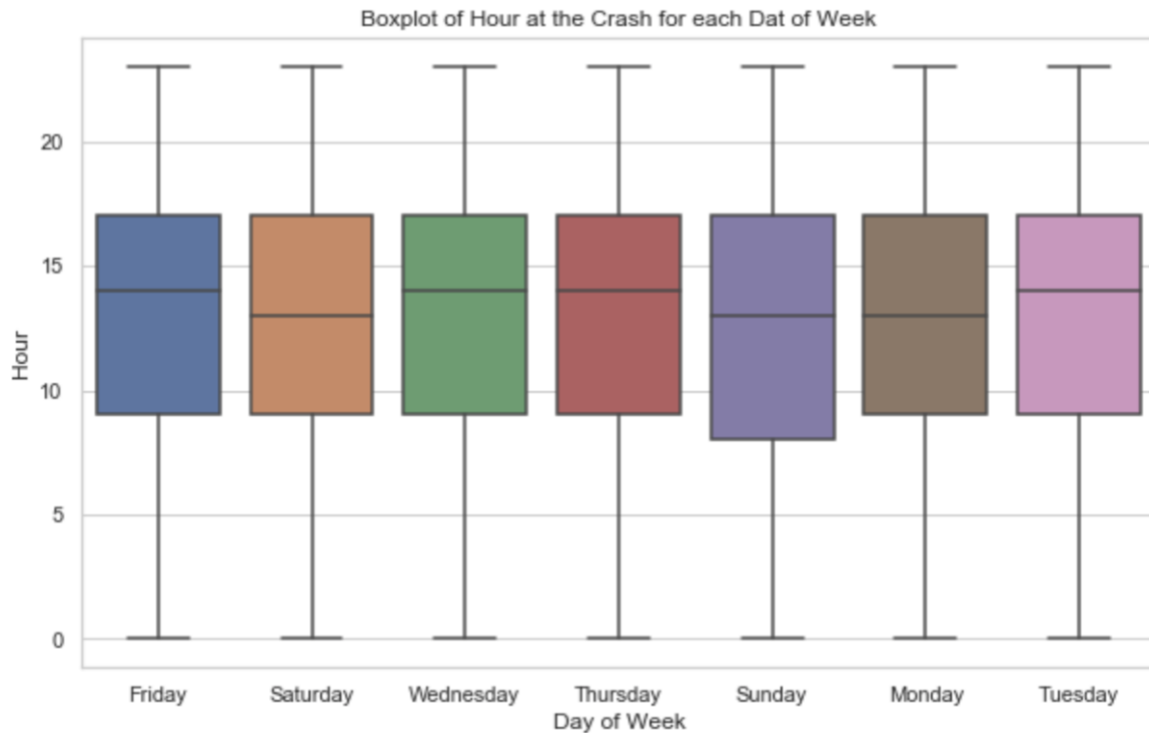
I want to know what time of a day involves most crashes, so I plotted a box plot. Its result is not surprising, with most crashes (25th percentile to 75th percentile) happening in the day from 9:30 to 17:00 for all 7 days of a week.

4. Scatterplot

Scatterplot of Number of Crashes in each Month and Year
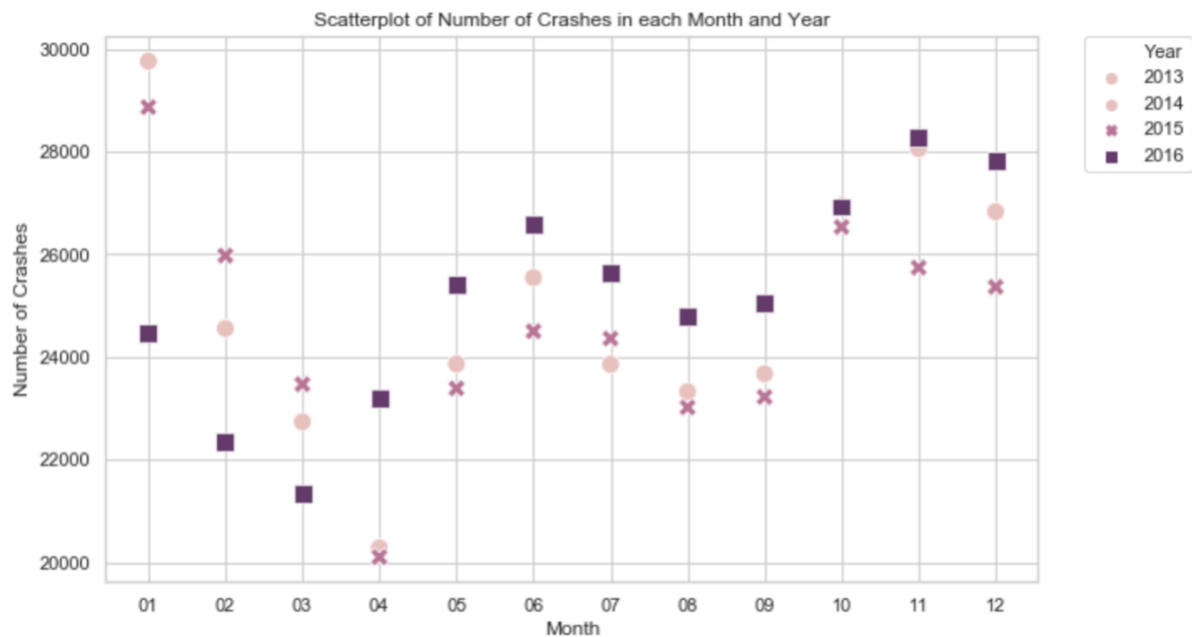
Which month has most crashes in a year? To know this, I plotted a scatter plot of number of crashes in each month for 3 years 2014 to 2016. Very interestingly, the trend is quite

similar in these 3 years, with most crashes happening in January, February, June, November and December.

5. Bubble Map

Bubble Map of Safety Equipment when Injury is Not Severe
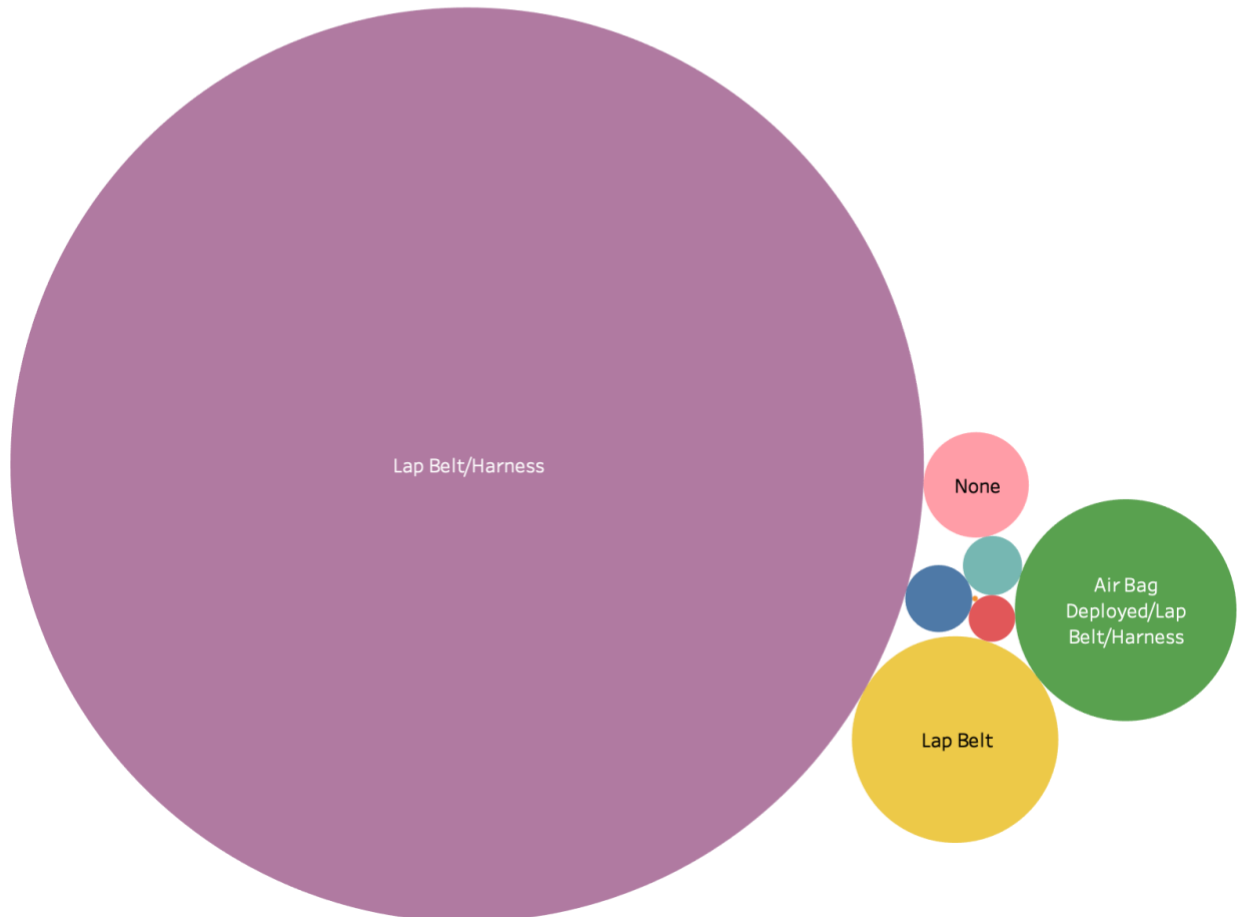
I filtered injury severity to be 'Not injured' and 'Minor', indicating an injury is not severe. Then I drew a bubble plot with Tableu to see the percentage of different safety equipment. Result shows that lap belt and air bag are quite useful in preventing severe injuries.

6. Chloropleth Map

Chloropleth Map of Average Completed Insurance Courses in Each County of New York
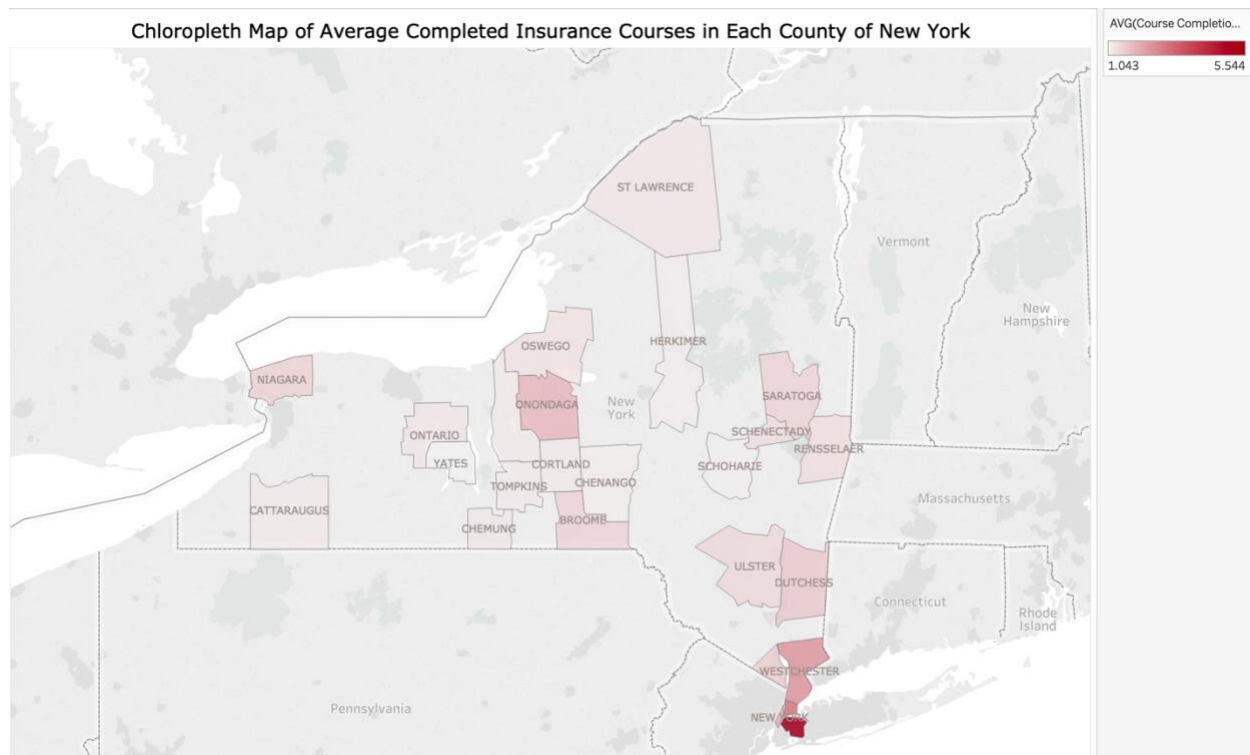
I'd like to know the insurance completion situation in the counties with crash records in New York State. So I plotted a chloropleth map for clearer illustration.

7. Heat map

Heat Map of Percentage of Crashes at each Injury Severity Level with different Vehicle Type
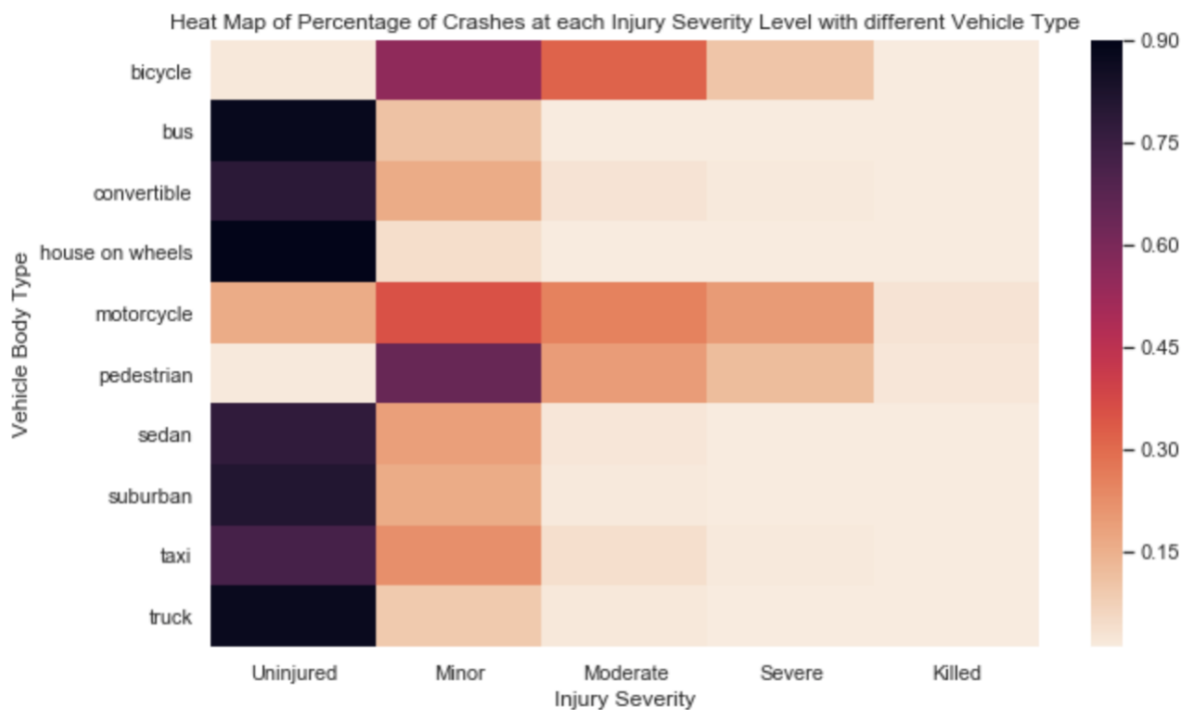
I'd like to know whether injury severity is correlated with vehicle type, so I made a heat map between these two. Results show that pedestrian, bicycle and motorcycle are

extremely dangerous, highly correlated to severe injuries. On the other hand, truck, house on wheels, bus and suburban, all with large bodies, are safer in a crash with lower rate of severe injuries.

8. Stacked area graph

Stacked Area Graph of Number of Crashes in each Month in Year 2015
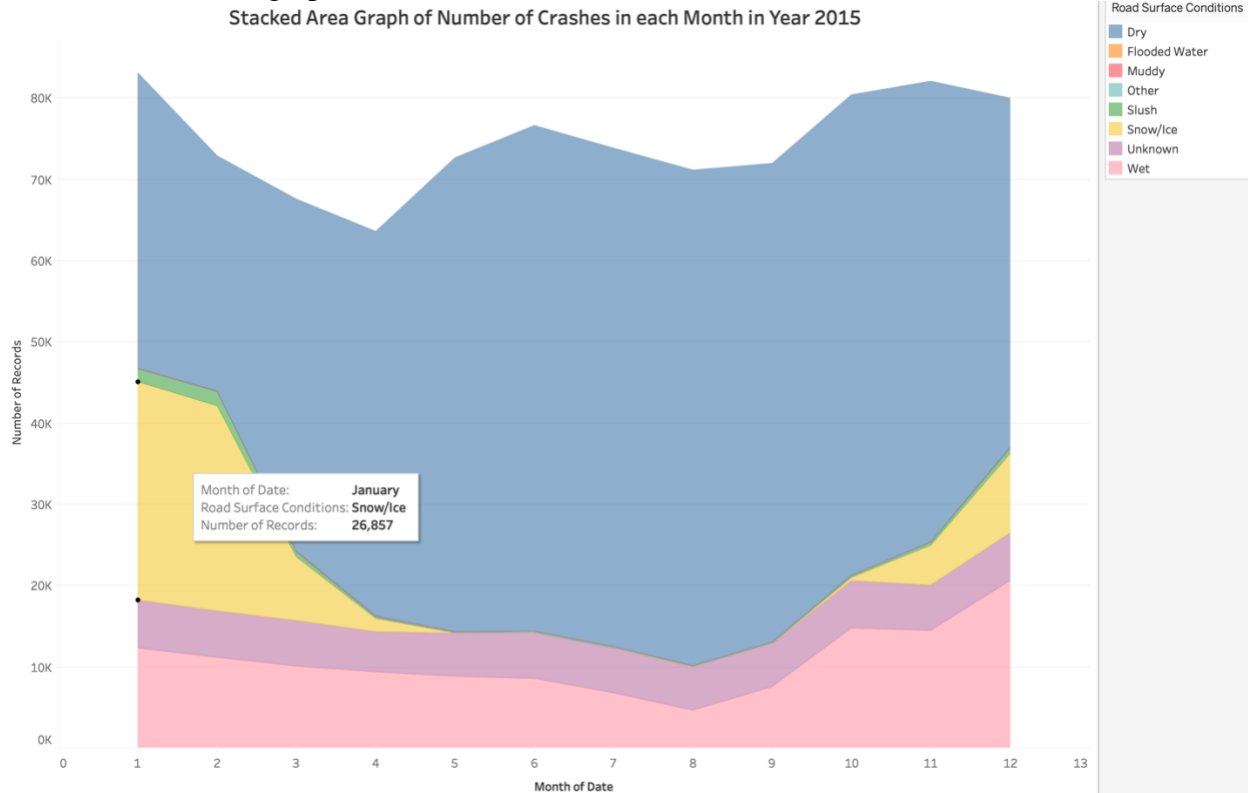
I plotted a stacked area graph to see what road conditions are involved with more crashes in 12 months in 2015. Results show that in the last two months and first 3 months of a year (when it's cold), snow and ice cause a lot of crashes compared to the others.

9. Treemapping

Interactive Treemapping of Numer of Crash Records at Different Road Types when Range of Number of Vehicles Involved Change

Number of Vehicles Inv...
13   50
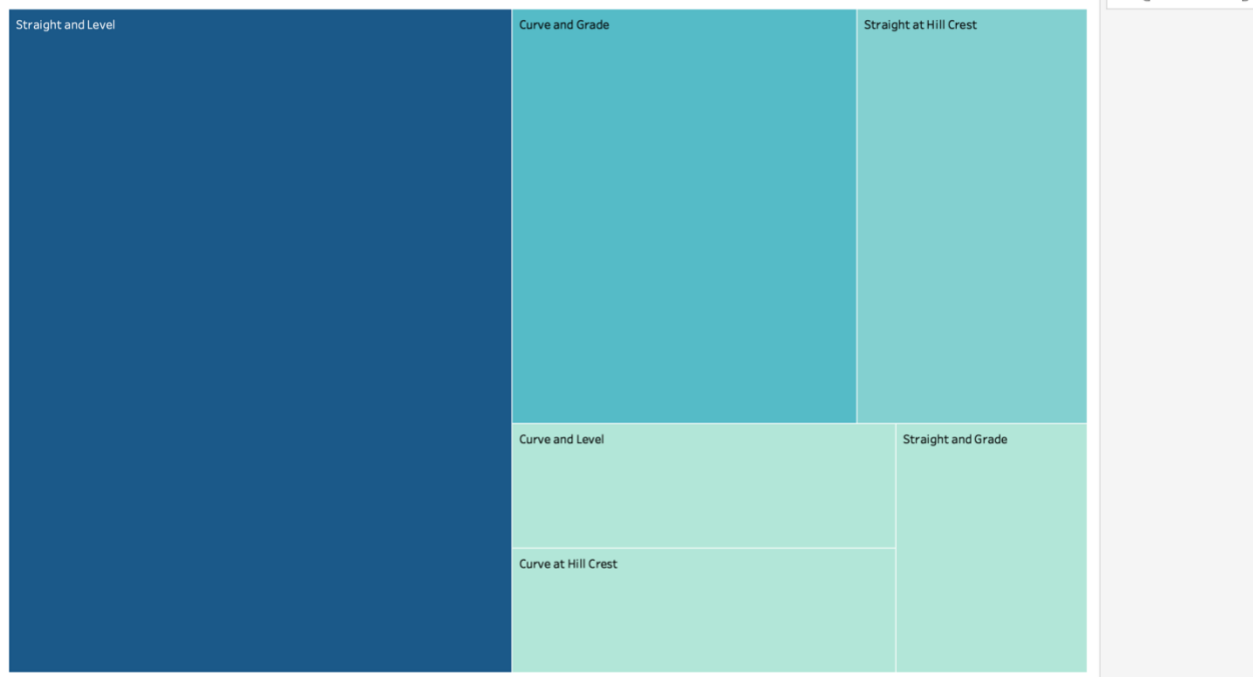
I made a treemap of number of crashes at different road types. I controlled range of number of vehicles involved in each crash by filtering, and find that when more vehicles are involved, higher percentage of roads are curve and grade instead of straight.

10. An interactive plot

Total Number of Motor Vehicle Crashes at Different Facilities from 2000 to 2015
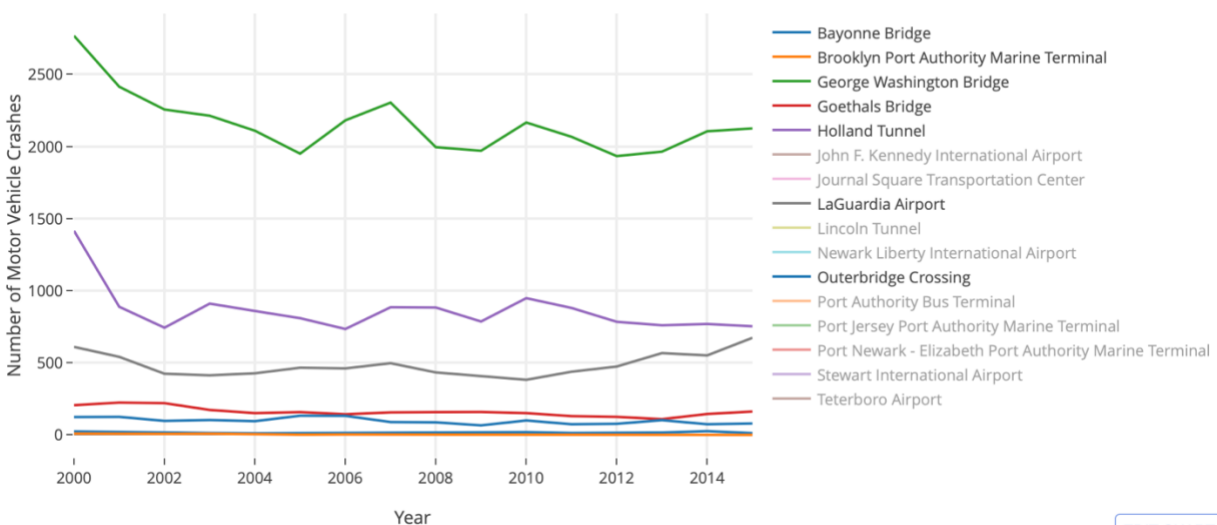
Finally, I made an interactive line plot with plotly, to see the number of vehicle crashes at different facilities (briges/crossing…) from 2000 to 2015. Results show that bridges involve more crashes compared to tunnel, airport or crossing, etc.

Your storyline:
Several factors can influence the safety of your road trip:
1. Sit in the back of a car can make you be injured more slightly in a crash.
2. Winter (from November to March next year) is much more dangerous because of the snow and ice increases probability of vehicle crashes a lot.
3. Lap belt, helmet and air bag are necessary equipment to keep you safer in a road trip.
4. Vehicles of larger bodies (house on wheels, trucks, buses, suburbans) are safer than smaller vehicles (pedestrian, bicycles, motorcycles).
5. Curves and grades are more likely to have big accidents (which involve more vehicles in one crash) compared to straight roads.
6. Bridges involve more crashes compared to tunnel, airport or crossing, etc.

Another thing to mention is that, children/baby are often in vehicle crashes but have no insurance. So better to insure your child since a very young age.

Results/Summary/Conclusion:
In conclusion, there are some suggestions for a safer road trip:
1. Sit in the back of a car.
2. Be careful in winters.
3. Obey the regulations to equip yourself with lap belt, helmet or air bags.
4. Choose larger vehicles, avoid motorcycles.
5. Pay special attention to curves or grades.
6. Be careful when driving through bridges.

Appendix Containing All Code:
Codes are uploaded in GitHub (https://github.com/ywang-hanna/visualization_final_project)

Link to your GitHub page with this analysis:
https://github.com/ywang-hanna/visualization_final_project

Citations:
Seaborn documentation: https://seaborn.pydata.org