0930Report–Codebook and Explorartory Analysis

Boyang Zhang

Introduction

The data used in this research is obtained from the Fatality Analysis Reporting System of the National Highway Traffic Safety Administration, focused on the Maryland and Washington, DC area from years 2011 to 2015. Other information, such as total population and total vehicle miles traveled, could be obtained from State Traffic Safety Information and other resources. Our main goal was to identify communities that might be at a higher risk of crashes in MD and DC area. If plausbile, we could broaden our analysis on national level, i.e. identify states that might be at a higher risk of crashes. Another option would be analyzing FARS dataset from crash level, vehicle level and person level respectively.

Accident file(2011-2015):

Among 2336 accidents and 2524 fatalities, several crash-level related factors are listed as follows. Here, we merged data based on common columns shared by accident data files from 2011 to 2015.

1. Time-varying variables

- Seasonality (month): there is no extreme difference across months
- Day of week: on Mondays and Sundays, accidents are more likely to happen in MD and DC area.

2. Geographical factors

- Intersection type: most of the accidents do not occur in intersection area and remaining accidents are mostly occurred in four-way intersection and T intersection.
- National highway system: among all accidents, a quarter of accidents are in NHS.
- Trafficway: identify specific roads

3. Enviornmental factors

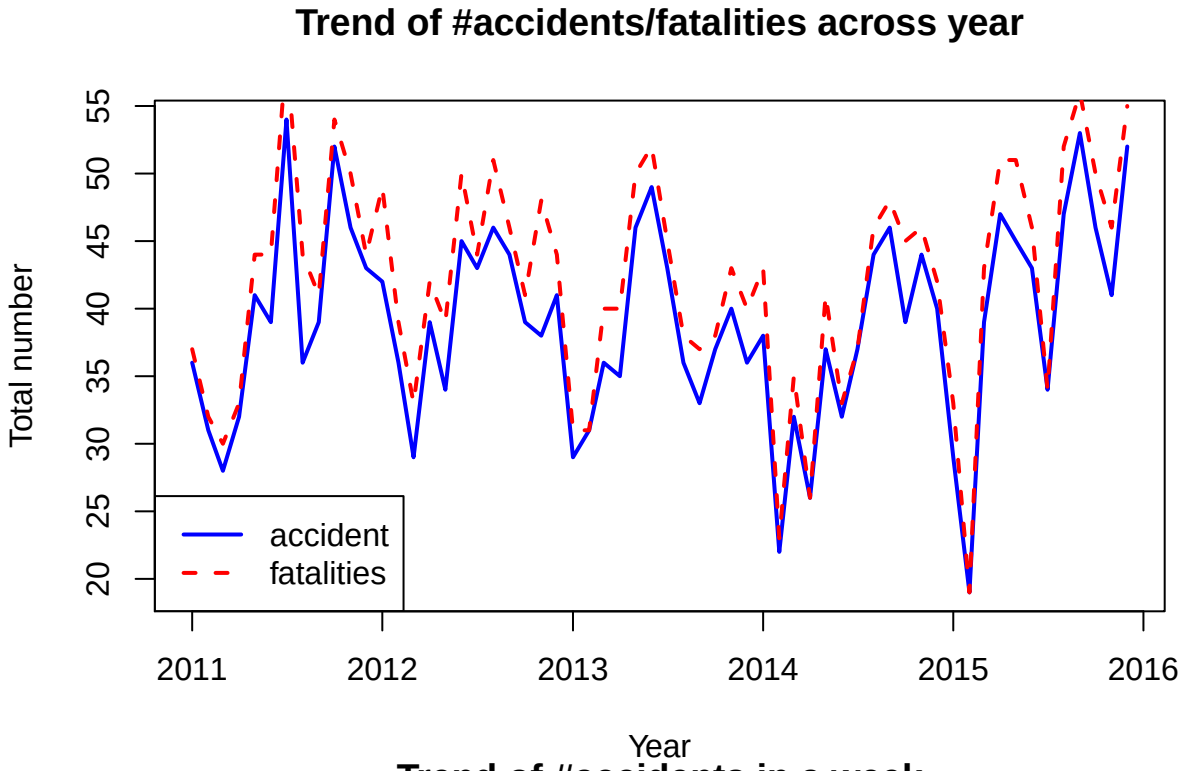
- Light condition: indicate type/level of light that existed at the time of the crash(day or night).
- Weather: summary of weather condition. Most of the accidents do not occur during adverse weather conditions.

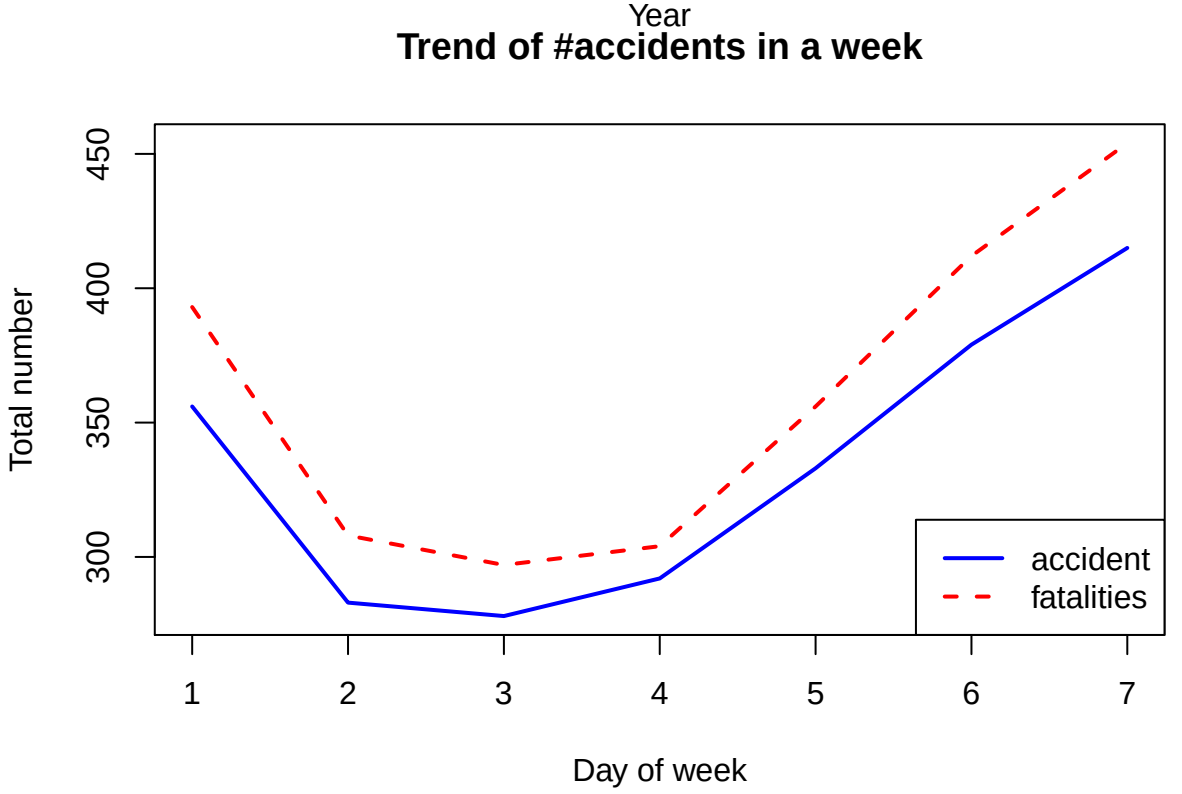
4. Information about crash

- Number of persons not in vehicles in transport: most of the accidents do not involve other person outside transporting vehicle, but still a quarter involves other people outside the car.
- Number of persons in transporting vehicle ranging from 1 to 18.
- First harmful events(accidents): first cause of crash and the top 5 harmful events are namely, vehicle in transport, pedestrian, tree, curb and guardrail face.
- Crash related factors: This contains too much missingness. Similar information can also be obtained from vehicle files and person files (vehicle-level related factors, driver-level related factors and person-level related factors in the Person data file).
- Drunk drivers: There are 737 drunk driving cases.

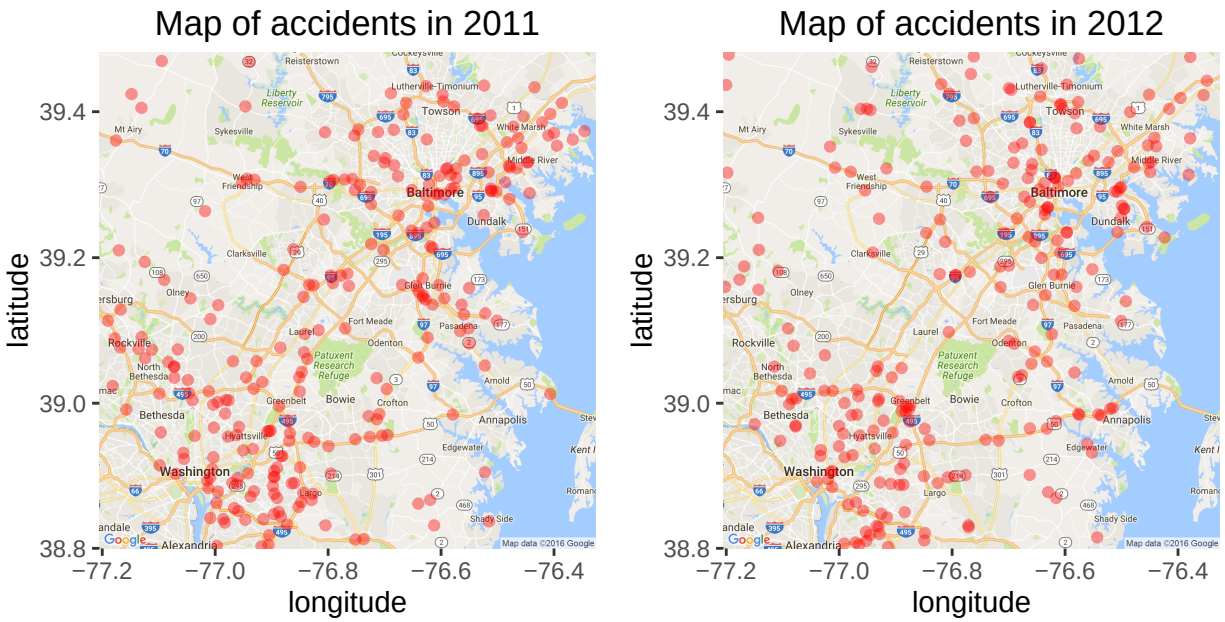
Explorartory analysis

Time varying factors





Geographical factors



Vehicle file and its related multiple related files(2011-2015):

Among 3554 vehicles involved in accidents during 2011 to 2015 period, several vehicle-level related factors are listed as follows. Here, we looked into its related multiple related files and found that 3 additional files could be useful for further analysis. Also, we merged data based on common columns shared by vehicle data files from 2011 to 2015.

1. Basic information

- Number of occupants in each vehicle
- Vehicle number: consecutive numbers of each accident, where 0 represents non-motorist in person file. It ranges 1-5, indicating that most adverse situation would involve 5 vehicles.
- Body type: classification of vehicle. Top five categories of accident-related cars for past 5 years are: 4-door sedan, compact utility, motorcycle, standard pickup and 2-door sedan/coupe.

2. Crash-related variables

- Hit and run: 145 hit and run cases for past 5 years.
- Extent of damage: most of the crashed cars lead to disabling damage, followed by functionl damage.
- Most harmful events(vehicle): top 3 accident-related harmful events to a vehicle are vehicle in transport, pedestrian and tree.
- Crash type based on first harmful events and previous circumstances (re-categorize into 6 categories): single driver, same trafficway opposite direction, changing trafficway, intersecting path and miscellaneous.
- Pre-existing vehicle defects: in Factor file (multiple responses)

3. Driver-related factors

- Driver's distraction: in Distract file (multiple responses)
- Driver's physical condition: in DrImpair file (multiple responses)
- Previous recorded crashes/suspension/revocation/speeding: needed to be recoded to 3 levels (i.e.,never,moderate,serious)

Person file and its related multiple responses files

Among 5781 people who are involved in accidents, the following are some important factors. Here, we merged data based on common columns shared by person data files from 2011 to 2015.

- Age/sex/race: demographic information.
- Injury severity: nearly half of people are fatally injured, followed by no injury, minor injury, serious injury.
- Seat position: for those who are in a vehicle, front seat is potential risk position.
- Air bag deployed: missingness
- Alchol/drug usage: missingness