Apart from some external environment factors, individual’s characteristics also are a significant factor to influence the likelihood of a crash. Therefore, the following analysis will pay more attention to some person-level data elements, separately focusing on person sex, age, and road user class.

1. **Person sex**

***Person sex*** (*p\_sex*) is a column shows the sex of the individual who is involved in a traffic accident. For convenience, all unknown and missing information variables (which include not applicable elements and those are not provided by jurisdiction) will be dropped.

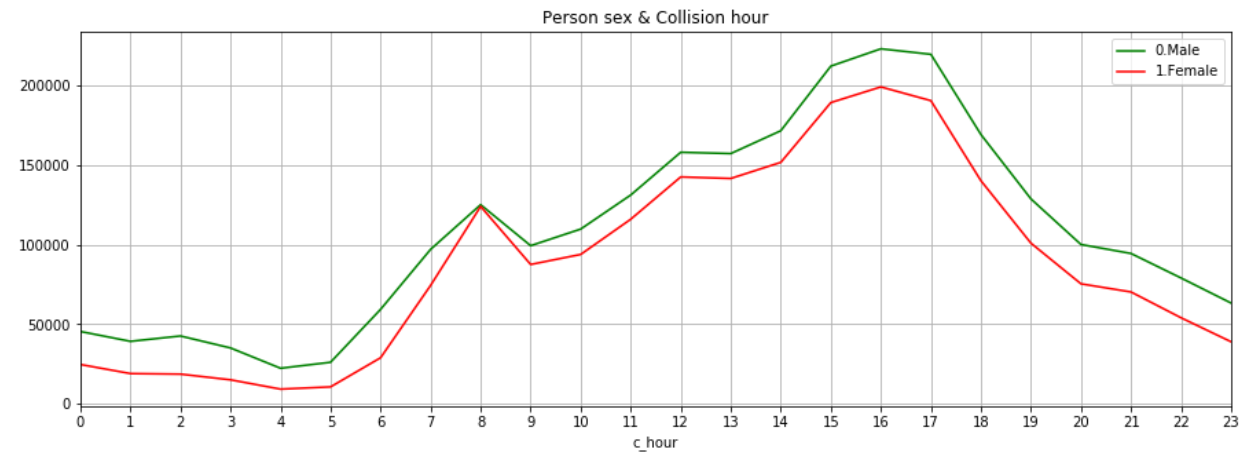
In this section, we will analysis three factors to person sex: collision hour; medical treatment required; collision configuration.

* 1. Person Sex & Collision Hour

***Collision Hour*** (*c\_hour*) is a column simply shows when the collision happened. Number from 0 to 23 correspond to 24 hours. For example, 0 indicates form midnight to 0:59; 14 represents from 14:00 to 14:59.

In the whole sample, we grouped the data by sex (*p\_sex*), and then using *count* method to account the number for each sex. So, data includes 2, 608, 529 men and 2,116, 502 women. Therefore, we might expect that man will more likely to involve in a collision.

In order to analyze the relation with collision hour, we need slice and group the data by sex (*p\_sex*) and collision hour column (*c\_hour*). Then using *size* and *unstack* method to aggregate the result. In the end, we plot a line chart with grid line for data visualization purpose:

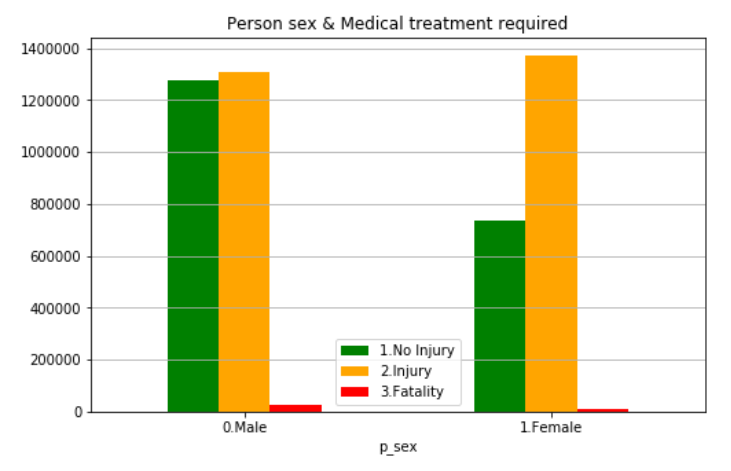


Our expectation is confirmed by the line chart, men are more likely to involve in a collision accident at all times. However, the collision times of male and female are very close to each other at 8 am, around 125047 and 123961 times. Also, for both male and female, the riskiest time period is form 3 pm to 5 pm, and then the risk has a monotonically decreasing trend.

* 1. Person Sex & Medical Treatment Required

***Medical Treatment Required*** (*p\_isev*) is a column indicating the degree of casualty. After dropping all unknown and missing information elements, dataset indicated three levels: *No injury*, *Injury* and *Fatality*.

By using the same method as before, we can easily aggregate the result into a two-level index table (p\_sex and p\_isev), and then plotting a bar chart.



If we focus on No injure and Injure now, we can conclude that female is more likely to injure in an accident, compared to male. In another word, women are more likely to have a serious accident. Plotted in this scale, it is hard to get information about fatality. Therefore, we will plot these same data on a log scale:



By comparing the red bars, we can see that men have greater likelihood to die in an accident. Numerically, men are twice as likely to die as woman.

* 1. Person Sex & Collision Configuration

***Collision Configuration*** (*c\_conf*) is a variable to describe the type of a traffic accident. Dataset includes 4 major categories, which are single vehicle in motion, two vehicles in motion with the same direction, two vehicles in motion with the different direction, and one vehicle hits a parked motor vehicle. Also, the specific type of collision is indicated within each major category. Therefore, more details of each of them will be provided as following.

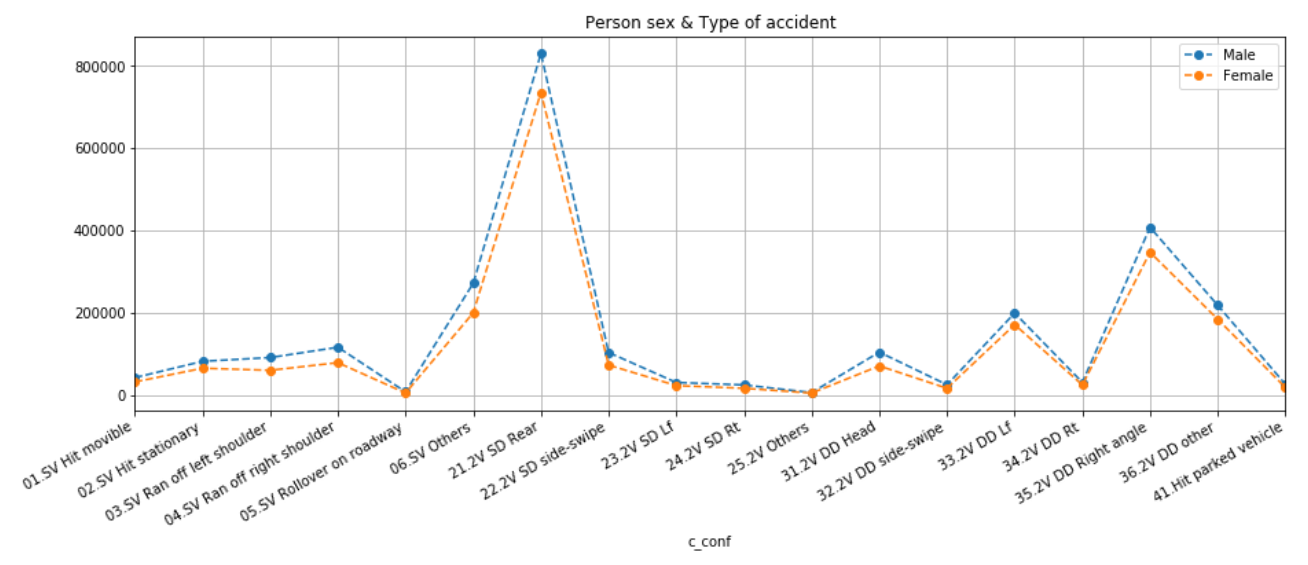
There are six types of collision when only one vehicle is involved. The first type is hitting a moving object, such as an animal. The second type is hitting a stationary object, like a tree. The third one is vehicle ran off left shoulder, including rollover in the left ditch. Corresponding with the previous one, the fourth one is it ran off right shoulder, which include rollover in the right ditch. Rollover on roadway is the fifth type, and the last one is others.

When two vehicles are moving in the same direction, five types of collision might occur. The first one is rear-end collision. The second one is side swipe, which means the sides of vehicles hit one another. The third type is one vehicle passing to the left of the other, or left turn conflict. The fourth one is exactly the opposite to the previous one. The last one is other types.

Six kinds of crashes might happen when two vehicles are moving in the different direction. The first one is head-on collision. The second type is approaching side-swipe. The third and fourth types are left turn or right turn across opposing traffic. The fifth one is right-angle collision. The last one is other types.

The final category only includes one type which two vehicles are involved in, but one hits a parked motor vehicle.

Now, the similar method will be used to analyze the relation between sex and collision configuration, which is shown as below.



It is obvious that male always has higher collision risk than female. Basically, male and female have a similar pattern among all kinds of collision configurations. Which means accident types do not have a significant sexual differentiation. Rear-end collision is the most frequent type of the traffic accidents for both man and woman, and Right-angle side collision is ranked as the second. Additionally, Rollover on roadway is the least common type of accident.

1. **Person age**

***Person age*** (*p\_age*) is numerical column, the age is from 0 to 99 years old. It is noteworthy that 0 means less then 1 years old and 99 represents 99 years or older. All other numbers, between 1 and 98, correspond to 1 to 98 years old. In doer to focus on generation differentiation instead of each age, we binned person age into 10 bins. For example, 10 represents the age from 11 to 20; 20 represents the age from 21 to 30 and so on. For analysis purpose, we still choose to drop all unknown and missing information data.

* 1. Person Age & Collision Hour

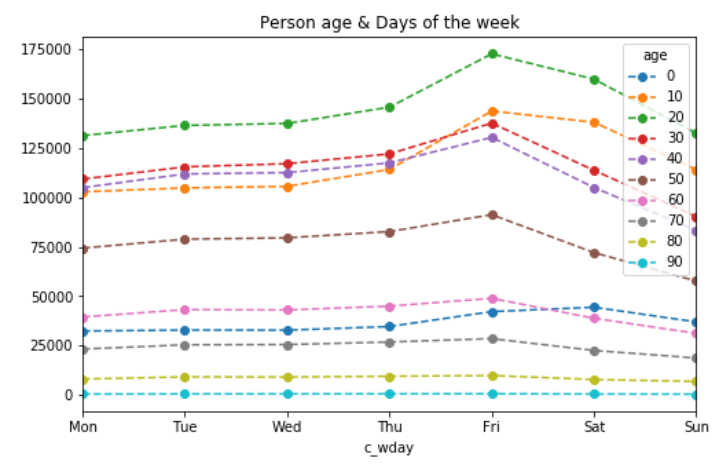
Like the sub-section 1.1, we define x-axis as collision hour in one day. Also, 10 different-color lines represent 10 age-groups.



The graph shows that the most dangerous population is from 21 to 30 years old. Also, the youth whose age is from 11 to 20 are more likely to involve in a collision at night, especially from 9 pm to 1 am. At morning (8 am to 11 am), three age-groups (21 to 30; 31 to 40; 41 to 50) has the similar number of accident. In the end, graph shows the working people share the same pattern, which is the rush hour are more likely to collide.

* 1. Person Age & Day of week

**Day of week** (*c\_wday*) is a variable, which contains a series of numbers from 1 to 7. Therefore, 1 to 7 means Monday to Sunday.

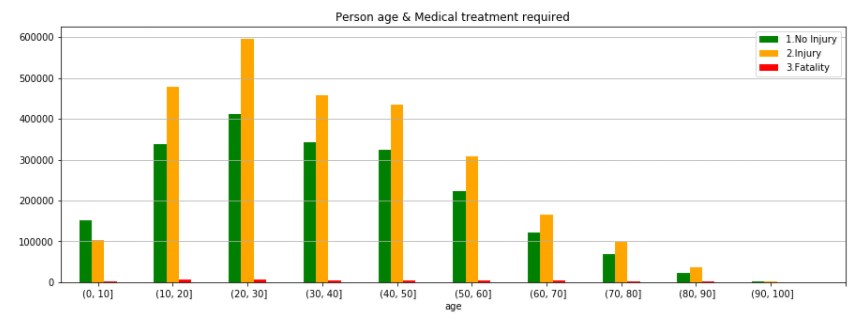


The plot shows that during the whole week, young adults (from 21 to 31) are the riskiest population. Risk decrease with age increases, for example the age from 41 to 50 are less likely to involve in an accident, compared with whose age is from 31 to 40. Also, in-accident probabilities of the youth are significantly increasing over the weekend. Which is realistic, because the most of teenagers or young drivers are still in school.

In addition, the most of age-groups, like the teenager, young adult and adult, point out most collisions happened on Friday. Also, teenagers and young adults involved collisions happened much less on Monday. In contrast, adults are less likely to involved in an accident on Sunday.

* 1. Person Age & Medical Treatment Required

Considering the normal scale first.



Except for whose age below 10, other age groups get the similar conclusion that people are more likely to injure in collision, and the likelihood decreases with age. Same as before, the graph shows the youth are the most dangerous group. The injuries of young adults are around twice of those of whose age between 51 to 60.

Now, changing to a log scale with the same data.

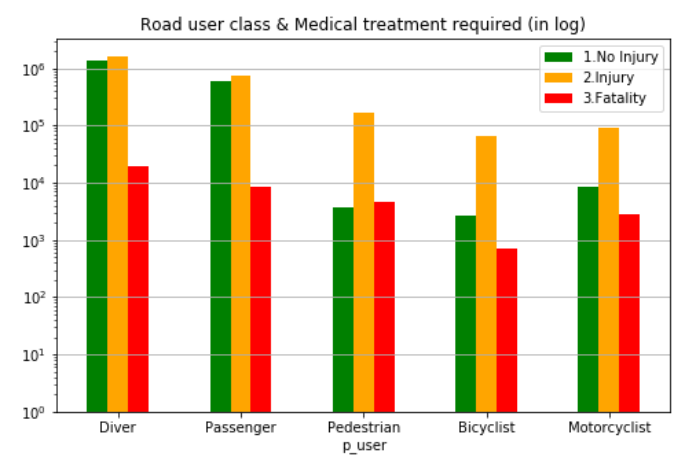
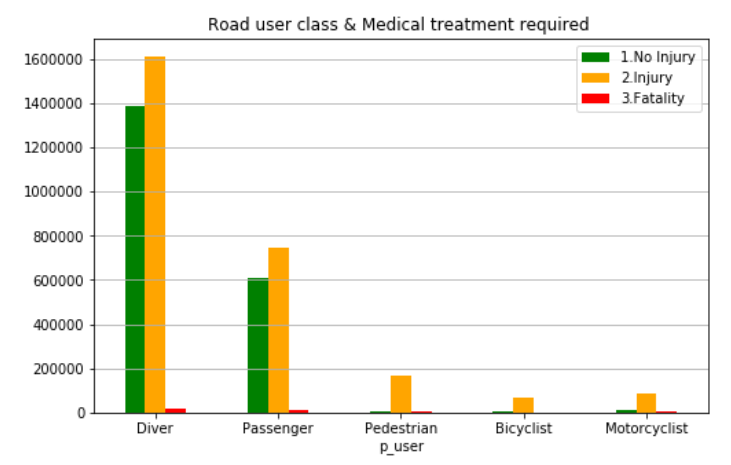


In contrast with the huge gap in injuries, young adults’ deaths are not outstanding, although it is the maximum. Additionally, people who above 90 are less likely to die in traffic casualty then all other age groups. Numerically, from 1999 to 2014, only 185 old people (above 90) died in an accident, compared with 7064 young adults died.

1. Road user class

**Road user class** (*p\_user*) shows when accident happened who is involved. Dataset provides 5 road users, with codes from 1 to 6. Thus, 1 is motor vehicle driver; 2 is motor vehicle passenger, 3 is pedestrian; 4 is Bicyclist; 5 is Motorcyclist; and the last one includes others and unknown. For consistency, the last elements will be dropped.

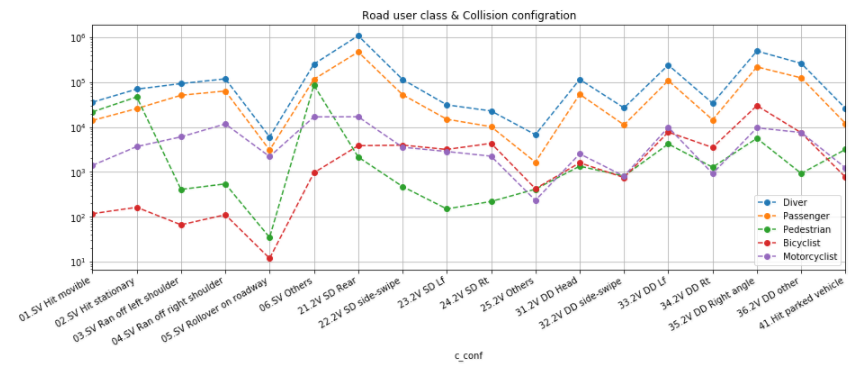
* 1. Road User & Medical Treatment Required



The major class involved in an accident is drivers, they are twice likely to get hurt or die than passengers. In particular, for each three degrees of casualty, the cases of driver are more than the sum of the rest of road users. Also, among all 5 categories, bicyclists are the safest, because of the lowest injuries and deaths.

Additionally, as far as drivers and passengers is concerned, not injury and injury separately account for half of total cases since fatality only accounts for about 0.06% in both road users. Whereas, a collision has above 85% possibility to cause injury when either pedestrians, bicyclists or motorcyclists are involved in.

* 1. Road User & Collision Configuration



In the end, it is no surprise that diver and passenger have the identical pattern, both of them are more likely to involve in a rear-end collision, and the second common type is right-angle collision. Single vehicle rollover on roadway causes pedestrian or bicyclists to injure or die is infrequent. However, pedestrians are more likely to involve in an any other single vehicle collision configuration; for bicyclists, they are more likely to get hurt in a right-angle collision. Among all types of accidents, Motorcyclists get a middling ranking.

**Conclusion:**

1. **Person Sex and Accident**

Based on previous analysis, some observations can be concluded as follow:

* Men are more likely to involve in a collision accident than women.
* Female is more likely to injure in an accident.
* However, men are more likely to involved in a fatal accident. Numerically, men are twice as likely to die as woman.
* Accident configurations and collision hour do not have a significant sexual differentiation.

1. **Person Age and Accident**

* The most dangerous population is from 21 to 30 years old, called young adults, and the in-accident probabilities decreases with age increases.
* The youth whose age is from 11 to 20 are more likely to involve in a collision at night, especially from 9 pm to 1 am.
* People who are in working age share the same tendency, which is the rush hour are more likely to collide.
* As far as the teenager, young adult and adult are concerned, most collisions happened on Friday.
* Also, in-accident probabilities of the youth are significantly increasing over the weekend. In contrast, adults are less likely to involved in an accident on Sunday.
* The injuries of young adults are around twice of those of whose age between 51 to 60. However, all cases show that fatality of young adults are not outstanding. Which means young adults are more likely to involved in a not fatal crash.

1. **Road User Class and Accident**

* The major road user class involved in an accident is drivers, they are twice likely to get hurt or die than passengers.
* Bicyclists are the safest, because of the lowest injuries and deaths.
* A collision has above 85% possibility to cause injury when either pedestrians, bicyclists or motorcyclists are involved in.
* Drivers and passengers are more likely to involve in a rear-end collision. Pedestrians are more likely to involve in other types of single vehicle collision. For bicyclists, they are more likely to get hurt in a right-angle collision.
* Also, single vehicle rollover on roadway causes pedestrian or bicyclists to injure or die is infrequent.