

## 1. The explored data set and application

The data set was obtained from the website of NHTSA (Nation Highway Traffic Safety Administration) [http://www-nrd.nhtsa.dot.gov/database/nrd-11/veh\\_db.html](http://www-nrd.nhtsa.dot.gov/database/nrd-11/veh_db.html). This set includes 5 tables, recording for every testing session the vehicle information, environment configuration and sensor readings. The actually used data set has **1961** instances, each of which is composed of **21** variables covering the above test configuration and measurements information.

My exploration focused on experiments of front crashing into barriers. The involved cars were made in 1995~2005, and all equipped with 3-point safe belts and airbags. Further, there were 1~6 dummies installed on cars and each of them were measured for several injury indicators, including HIC (Head Injury Criterion), CSI(Chest Severity Index) and L/RFEM( Left/Right Femur Peak Load Measurement).

HCI ( Hierarchical Clustering Explorer) was employed to explore the relation between variables. After the first puzzling hours in using it, I got more and more familiar with it and really like it. Following are five interesting phenomena I found in the exploration.

## 2. Findings

### 2.1. Heavier cars are more dangerous

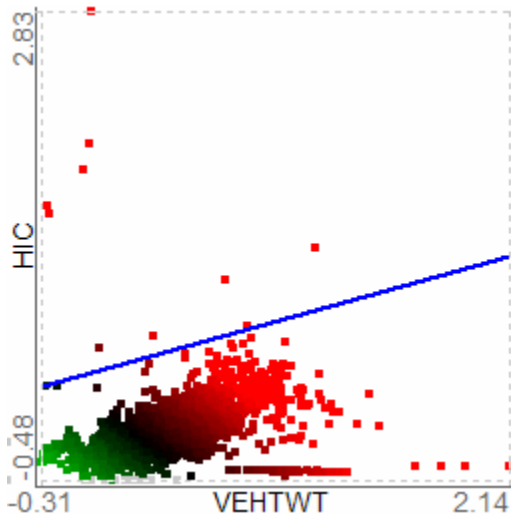


Fig1. Vehicle Weight - HIC

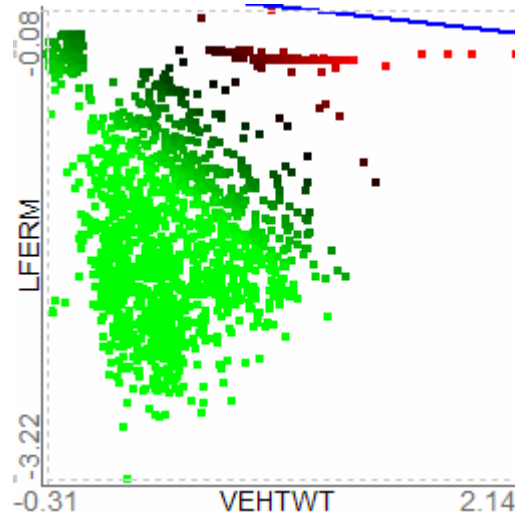


Fig2. Vehicle Weight – Left Femur

Figure 1 and 2 illustrates the correlation between vehicle weight and resultant injury on human head and femurs ( the figure for right femur is intentionally ignored because it is almost the same as the figure 2 ). An interesting thing is that the two body parts have different, if not completely reverse, response to increasing vehicle weight. Figure 1 clearly shows a trend that the heavier a car the more severe the head injury, but figure 2 tells a different story that the heavier cars come up with less head femur injury ( see the right half of the figure ).

## Application Assignment of CMSC 838S -- Chunyuan Liao

Despite the interesting discrepancy, we still argue that heavier cars are quite probably more dangerous, because the head injury is generally more fatal than femur injury.

For a deeper insight in head injury and femur injury, I stepped forward to examined the correlation between them, and got the following conclusion.

### 2.2. Head and chest usually get injured at the same time, but femurs may escape that

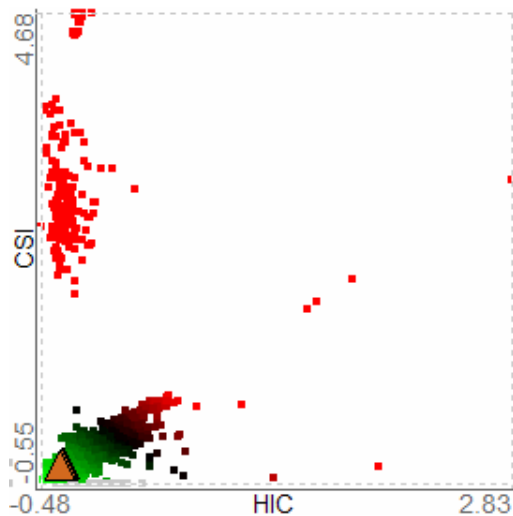


Fig3. HIC – Left Femur

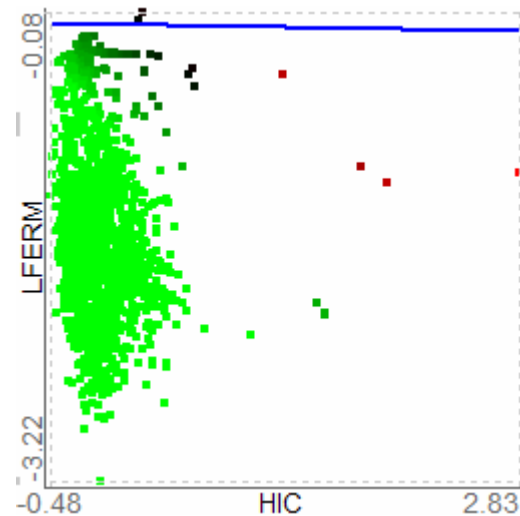


Fig4. HIC - CSI

Figure 3 indicate a positive linear correlation between HIC and CSI ( the red points on the left side are missing data, so ignored), while figure 4 shows that human femurs may get less harm on sever head injury. A possible explanation is that when the car crashed into the barrier, the human body actually began rotating due to such force by safe belts that the head went forward and the legs backward relatively. And the quicker the head moved ahead, the faster the legs withdrew. Consequently, femurs may get less harm than head, especially at a high speed.

### 2.3. Newer cars are longer/heavier, but not necessary safer

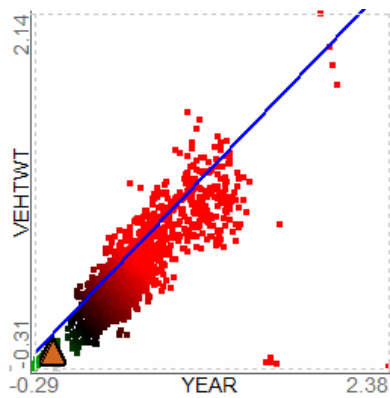


Fig5. Year – Vehicle Weight

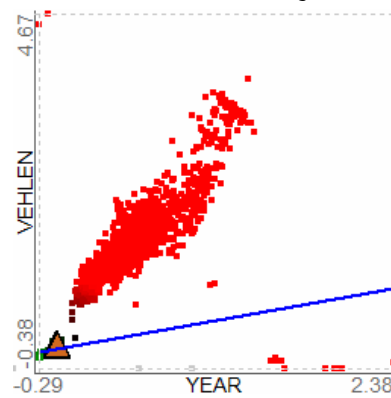


Fig6. Year – Vehicle Length

## Application Assignment of CMSC 838S -- Chunyuan Liao

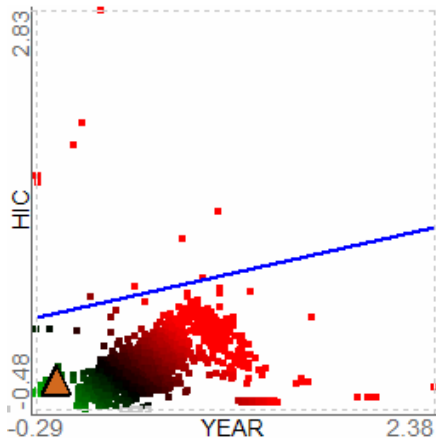


Fig7. Year - HIC

With figure 5 and 6, it is apparent that more heavy and long cars are used for crash testing than before. Assuming NHTSA's test is reasonably comprehensive and representative, we argue that it is a trend that people like heavier and longer cars now. But, unfortunately, it seems that the higher HIC also come up with new cars.

### 2.4. Which car is safer?

I tried "Profile Search" to find out which car is safer with the similar weather, road condition, speed and occupant position. By safer, we mean "lower HIC", since head injure is the most severe harm to human body.

Specifically, I set the speed to about 56m/h, occupant position to "driver seat", road condition to "dry" and surface type to "concrete". In HCE, I chose "Thresholds=1.000~1.000", "Search Methods=Model-based" and "Distance Measure=Pearson's r". Figure 8 shows the pattern I defined and the silhouette of the research results.

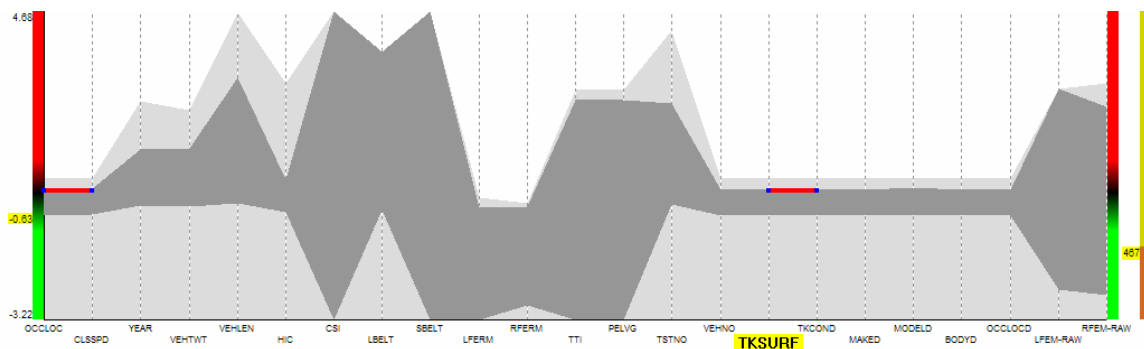


Fig8. Profile Search result

From the "Table View", I got the following top and least 3 safe cars in terms of HIC.

## Application Assignment of CMSC 838S -- Chunyuan Liao

Top safe	Make	Model	Year	Type	HIC
1	HONDA	ACCORD	2001	FOUR DOOR SEDAN	61.00
2	NISSAN	ALTIMA	2002	FOUR DOOR SEDAN	75.00
3	DODGE	INTREPID	1999	FOUR DOOR SEDAN	95.00

Least safe	Make	Model	Year	Type	HIC
1	DODGE	RAM WAGON VAN	1995	VAN	1474.00
2	TOYOTA	TACOMA	1997	EXTENDED CAB PICKUP	1411.00
3	DODGE	NEON	2002	FOUR DOOR SEDAN	1246.00

### 2.5. The crash test becomes tougher and tougher!

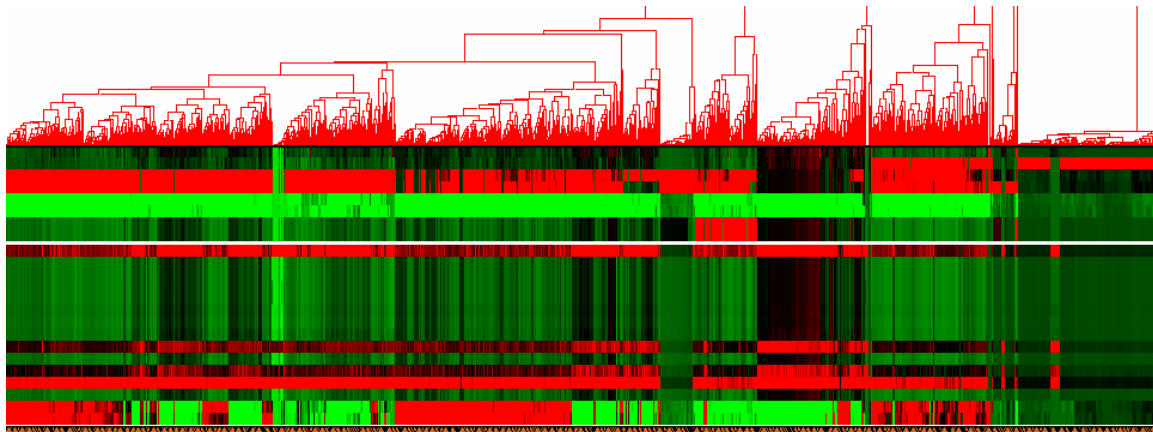
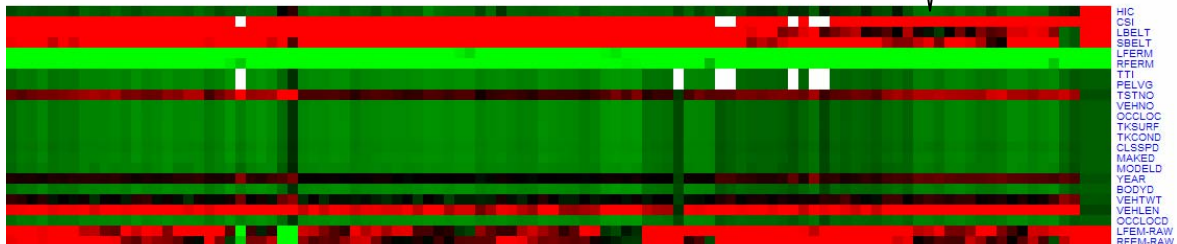


Fig9. Result Clustering based on injury



Cluster 1 : tests with mild injure



Cluster 2 : tests with severe injure

## **Application Assignment of CMSC 838S -- Chunyuan Liao**

I clustered the records based on indicator variables for human body injury. The most right cluster (cluster 1) stood out in respect of mild injury. Compared with other clusters, for example cluster 2, the test speed, vehicle weight, vehicle length, product year are all smaller. It turns out the less mild tests are conducted in these days.