The Application of Fatality Analysis Reporting System Data on the Road Safety Education of U.S., D.C., and PR Minors

SAS Student Symposium Competition
Fall 2015

SAS Student Symposium Paper

Team Name: M&M&M

I. Introduction

All public schools in the United States require health and safety education for their students. Furthermore, almost all states require driver education before minors can obtain a driver's license. Through extensive analysis of the Fatality Analysis Reporting System data, we have concluded that from 2011-2013 an average of 12.1% of all individuals killed in a motor vehicle accident in the United States, District of Columbia, and Puerto Rico were minors (18 years or younger). Our goal is to offer insight within our analysis in order to better road safety education to prevent future premature deaths involving motor vehicles.

II. Data

The data we used to analyze deaths due to car accidents was retrieved from the U.S. Department of Transportation, and the National Highway Traffic Safety Administration. It is important to note that the variable attribute descriptions changed after 2010, therefore in order to compare years prior to 2010 to those after, one must make note of these differences. It is for this reason that we limited our data analysis to the accident data in years 2011, 2012, and 2013.

Also important to note the original data source comes from a compilation of Police accident reports, death certificates, state vehicle registration files, medical examiner reports, hospital medical reports, and emergency medical services reports. This insures that our data is comprehensive and authoritative.

III. Problem Identification and Explanation

Following our research and understanding of the Fatality Analysis Reporting System data, we were able to identify a problem which offered the opportunity for an initial investigation into a profoundly effective solution based on our analysis. Currently, educational strategies for minor road safety are standardized. Thus, the proper education methodology of communication is not always portrayed to the appropriate target audience. According to the National Highway Traffic Safety Administration website, education for youth road safety lacks differentiation. For example, teen driver education is structured to be taught through the parents of teen drivers rather than the drivers themselves. Also, parent road safety education for young children (12 years or younger) is heavily focused on car seats while ignoring proper road safety practices to

teach young children.

We have concluded, in order for road safety education to be impactful, it is imperative to create custom lesson plans based on the specific educational variables that relate to the appropriate targeted audience. Our main objective was to create a basic understanding of what these essential educational variables were to offer an opportunity for a more effective and differentiated road safety educational system. In order to guide our analysis to a valuable conclusion, we posed the following question: *what common characteristics of minors involved in fatal car crashes can be used to improve minor education on road safety in the United States, District of Columbia, and Puerto Rico.*

IV. Data Cleaning and Validation

In order to prepare the dataset for answering the questions about accidents involving minors, the team evaluated which tables would provide valuable conclusions into necessary educational variables for differentiated road safety education. Tables NMCrash, NMImpair, NMPrior, and SafetyEQ were joined together on the variable ST_CASE. From there, the Person table from each respective year was used to join ST_CASEs and Person Numbers (this allowed proper person data to match an individual rather the entire accident) into the Non-Motorist tables (e.g., demographic information). This was done for years 2011, 2012, and 2013, and split if variable DEATH_YR=8888 (death year did not occur/still alive) or DEATH_YR=(an actual year, ex: 2012) to create Deceased Non-Motorists and Survived Non-Motorists tables for each year. These files were appended together to have one Deceased Non-Motorists Table (‘11-’13) and one Survived Non-Motorists Table (‘11-’13). For visuals on the variables used in the Non-Motorist table, *see Figure A.*

The Person tables from each respective year were used to build the In-Vehicle tables. The three were appended together to create a person table from ‘11-’13, and then split out on DEATH_YR (see above) to divide deceased and survived, and filtered on Person Type to include only In-Vehicle minors. These conditions provided the team with a Deceased In-Vehicle table and a Survived In-Vehicle Table. For visuals on the variables used in the In-Vehicle table, *see Figure B.*

For a full visual representation of the data clean process, s*ee Figure C.*

V. Analytical Procedures, Segment Analysis, and Profiling

Once the four files were prepared in Enterprise Guide, the subsets were brought into Enterprise Miner to run a SAS generated cluster analysis. Each subset was clustered individually, resulting in 4 independent clustered datasets, however similar restrictions and techniques were used across all four files. The number of clusters was automatically specified by the team, with an upper limit of 20 clusters, a minimum of 2, and a CCC cutoff of 3. All variables that were selected into the files from Enterprise Guide, with the exception of identification variables, were utilized in the clustering of the data points. The clustering method amongst all four groups was the Ward method, and the Internal Standardization setting was kept at Standardization.

After slight adjustments to the clustering criteria specific to each group, Segment Profile nodes were run on each of the four groups to gain a better understanding of the variables influencing each segment and how they compare to the mean across the data set. SAS standard profiling procedures were not adjusted for any of the groups (e.g., number of midpoints, minimum worth, etc.) as the team was satisfied with the profiling of the clusters standing alone. Because of the quantity of information available, the group only evaluated segments from any group that were greater than or equal to 10% of its respective data set (i.e., a segment from the Deceased Non-Motorist table that accounted for 40% of the set would be examined, but not 6%). From there, segments that were examined that provided actionable insights were unpackaged even further in order to narrow the scope of the analysis to a manageable degree. Under this evaluation criteria, 10 segments from the initial minor data set will be highlighted here and used as the basis of our recommendations for bettering road safety education.

I. Deceased Non-Motorists

The Deceased Non-Motorists group provided four clusters for evaluation (over 10% of the data), two of which provided actionable insights. These two segments were heavily influenced by the age variable, where one segment had all minors under 11 years old, and the other with all minors over 9 years old. In both cases, approximately 95% of the time these non-

motorist minors were not wearing safety equipment, which raised the question to our team about how can parents of young children and adolescent teenagers be incentivized to use more safety equipment (vests, helmets, etc.), and would that improve the mortality rate among these two profiles? For visuals on these variables and the Deceased Non-Motorist group, see *Figures D-F*.

II. Survived Non-Motorists

The contrary group in the Non-Motorist section of the analysis are those who were involved in the accident but survived. Again, this clustering process provided four segments that met evaluative criteria (as stated above), two of which provided real value in answering the proposed question. These two profiles again split on age (*see Figures H and I*). However, the older profile data showed that over 50% of the segment's data points were on a roadway that was not marked as a crosswalk appropriate space for a pedestrian to be standing. Additionally, the younger profile showed 67% of the crash times happened during rush hours (approx. 6:30-8:30am and 5:00-7:00pm), indicating that these younger children were on the roads (most likely accompanied by a parent) either immediately before or after the work day began (s*ee Figure H*). Both of these insights gave the team great knowledge about the intricacies of these fatalities and how they can be used to help better educate the youth on road safety measures. For visualizations on the survived non-motorist group, *see Figures G-H*.

III. Deceased In-Vehicle

The group that generated the most meaningful segments was the Deceased In-Vehicle group (either the driver or a passenger when the vehicle crashed and died). Four segments met the evaluative criteria, and four provided meaningful insight to the analysis. These profiles were divided on the variable "First Harmful Event", where two segments had events related to colliding with another vehicle in transport (labeled "Collision") and events related to swerving off of the road and/or hitting a tree, embankment, sidewalk, etc. (labeled "Independent Vehicle"). Collision groups (*see Figures J-L*), were also separated by driver vs. passenger, and seating positions (front seats, backseats, etc.) These shed light on whether the driver was aiming to protect themselves (in one segment 80% of the deaths were passengers) versus if the collision was head on (would impact the driver more so, one segment has the majority of deaths as drivers). From the independent vehicle groups, one profile indicated the driver was alone in the car the majority of the time when swerving off of the road, and the other profile indicated the majority of independent vehicle crashes happened later at night, where it may be easier to drive

off of the road. For visuals on the Deceased In-Vehicle groups, *see Figures J-N*.

IV. Survived In-Vehicle

The fourth and final group, the Survived In-Vehicle data points, had two profiles meet evaluative criteria, and while there was moderate insight conveyed from their data, the team did not have much of choice but to evaluate both profiles because they accounted for nearly 99% of the data in the file. One profile is contains 72% of the data set (*see Figure O*), and is nearly entirely passengers (*See Figure P*). Over half of these surviving passengers were in the backseat. The second profile was smaller, and was closer to a split on drivers and passengers. This set was more difficult to gain insight from because of the potentially skewness of the passengers to driver ratio (could range from 2:1 to 7:1). Visualizations for the Surviving In-Vehicle group are provided in *Figures O-Q*.

After examining the segments that we found meaningful (10 total), we assigned a face and a name to each of the profiles created (*see Figure R*). This was more of a visualization tool more so than an analytical one, however it should help the viewers personify the profiles that are presented. Please see *Figure S* for a breakdown of the different groups of data that were used, and the profiles that are associated with them. Additionally, see *Figure T* for an organizational chart to visualize exactly how the profiles originated from our Minors data set.

VI. Suggestions for Future Studies

Our team believes that this project warrants further study. It is apparent that there are gaps in the education of U.S., D.C., and PR citizens, which result in the premature death of minor children via car accidents. With a continuing study, we would be curious to have access to data of car accidents in the U.S., D.C., and PR where a death did not occur. It is our belief that with this data we would be able to perform predictive analysis that could lead to a model that would predict minor-death based on particular characteristics. While our current analysis is quite insightful, it does not allow us to perform predictive modeling, thus limiting our ultimate conclusions.

VII. Conclusion

Following a comprehensive review of our results, our team concluded that there is substantial evidence to suggest that there are valid contributing factors to the death of minor children due to

car accidents. We identified several significant variables from our profiles including lack of safety equipment, teen drivers, time of day, and off-road obstacles. It is our final suggestion that school health and safety programs attempt to expand road safety programs to target issues that face specific age groups. We also suggest that parents be compelled to receive road-safety training for their young children. Without further education on road safety for both drivers and pedestrians, it is impossible that the minor fatality rate will decrease over time.

VIII. Appendix

Figure A.

Select Data | Filter Data | Sort Data

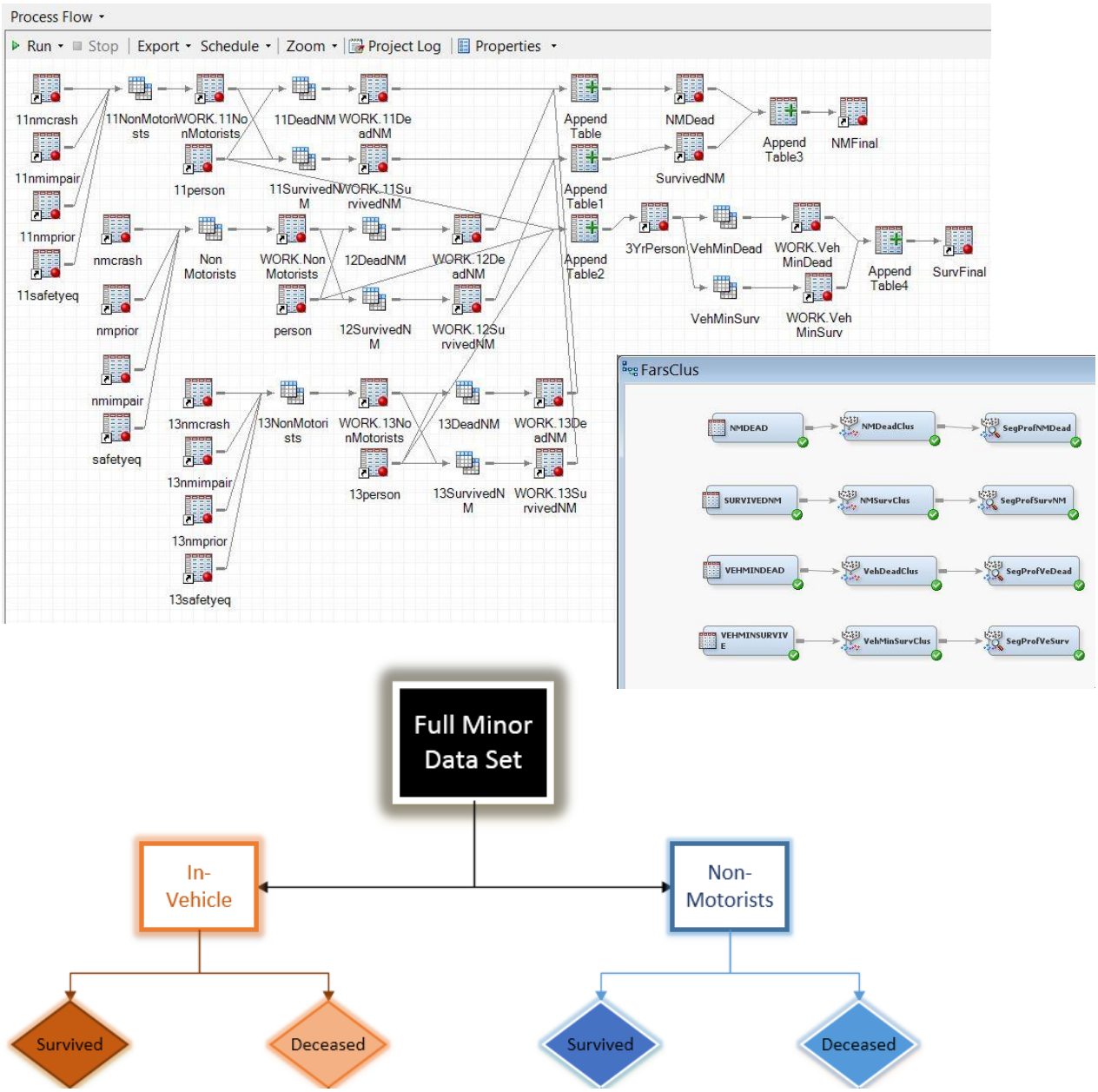
Column Name	Source Column
STATE (State Number)	t1.STATE
ST_CASE (Consecutive Number)	t1.ST_CASE
VEH_NO (Vehicle Number)	t1.VEH_NO
PER_NO (Person Number)	t1.PER_NO
ImproperAction (Non Motorist Action/Circumstances at Time of ...	t1.ImproperAction
Impairment (Condition (Impairment) at Time of Crash)	t1.Impairment
PriorAction (Non Motorist Action/Circumstances Prior to Crash)	t1.PriorAction
SafetyEquipment (Non-Motorist Safety Equipment Use)	t1.SafetyEquipm...
AGE (Age)	t2.AGE
SEX (Sex)	t2.SEX
RACE (Race)	t2.RACE
LOCATION (Non-Motorist Location at time of crash)	t2.LOCATION
DRUGS (Police Reported Drug Involvement)	t2.DRUGS
DRINKING (Police Reported Alcohol Involvement)	t2.DRINKING
DAY (Crash Date (DAY))	t2.DAY
MONTH (Crash Date (MONTH))	t2.MONTH
HOUR (Crash Time (HOUR))	t2.HOUR
MINUTE (Crash Time (MINUTE))	t2.MINUTE

Figure B.

Select Data | Filter Data | Sort Data

Column Name	Source Column	Su
STATE (State Number)	t1.STATE	
ST_CASE (Consecutive Number)	t1.ST_CASE	
VEH_NO (Vehicle Number)	t1.VEH_NO	
PER_NO (Person Number)	t1.PER_NO	
HOUR (Crash Time (HOUR))	t1.HOUR	
ROAD_FNC (Roadway Function Class)	t1.ROAD_FNC	
HARM_EV (First Harmful Event)	t1.HARM_EV	
SCH_BUS (School Bus Related)	t1.SCH_BUS	
AGE (Age)	t1.AGE	
SEX (Sex)	t1.SEX	
PER_TYP (Person Type)	t1.PER_TYP	
SEAT_POS (Seating Position)	t1.SEAT_POS	
REST_USE (Restraint System/Helmet Use)	t1.REST_USE	
REST_MIS (Any Indication of Mis-Use of Restraint System/Hel...	t1.REST_MIS	
DRINKING (Police Reported Alcohol Involvement)	t1.DRINKING	
DRUGS (Police Reported Drug Involvement)	t1.DRUGS	
RACE (Race)	t1.RACE	
LOCATION (Non-Motorist Location at time of crash)	t1.LOCATION	

Figure C.


Process Flow
Run
Stop
Export
Schedule
Zoom
Project Log
Properties
11nmcrash
11nmimpair
11nmprior
11safetyeq
11NonMotorists
WORK.11NonMotorists
11person
11DeadNM
WORK.11DeadNM
11SurvivedNM
WORK.11SurvivedNM
Append Table
Append Table1
Append Table2
NMDead
SurvivedNM
Append Table3
NMFinal
nmcrash
nmprior
nmimpair
safetyeq
Non Motorists
WORK.Non Motorists
person
12DeadNM
12SurvivedNM
WORK.12DeadNM
WORK.12SurvivedNM
3YrPerson
VehMinDead
WORK.VehMinDead
VehMinSurv
WORK.VehMinSurv
Append Table4
SurvFinal
13nmcrash
13nmimpair
13nmprior
13safetyeq
13NonMotorists
WORK.13NonMotorists
13person
13DeadNM
13SurvivedNM
WORK.13DeadNM
WORK.13SurvivedNM
FarsClus
NMDEAD
NMDeadClus
SegProfNMDead
SURVIVEDNM
NMSurvClus
SegProfSurvNM
VEHMINDEAD
VehDeadClus
SegProfVeDead
VEHMINSURVIVE
VehMinSurvClus
SegProfVeSurv

Full Minor Data Set
In-Vehicle
Non-Motorists
Survived
Deceased
Survived
Deceased

Figure D.

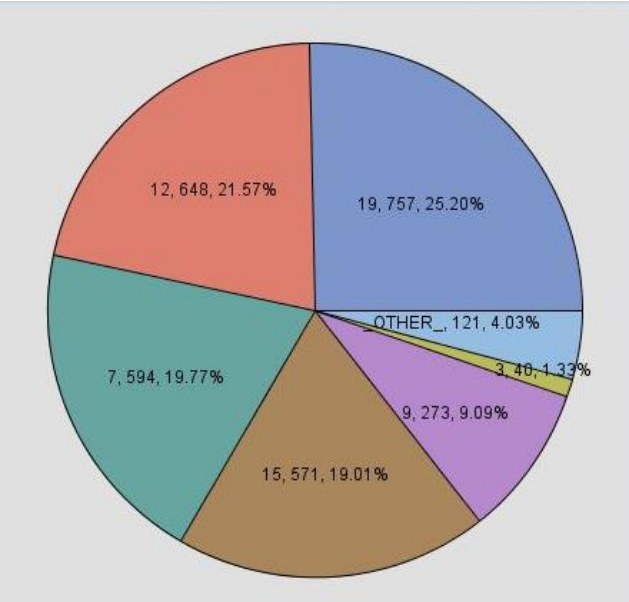
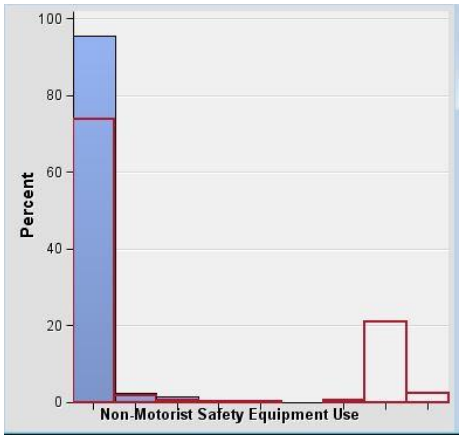


Figure E.



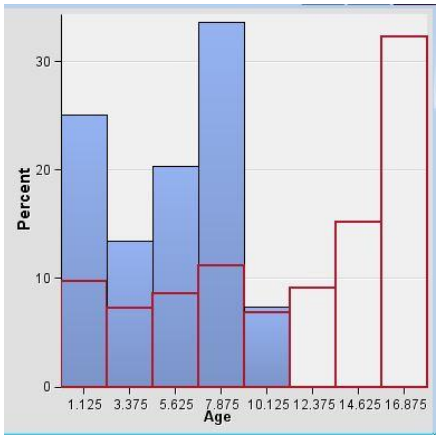


Figure F.

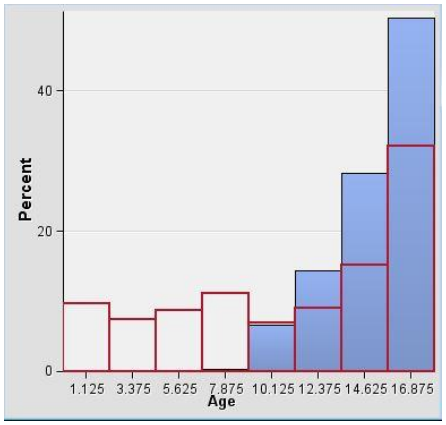


Figure G.

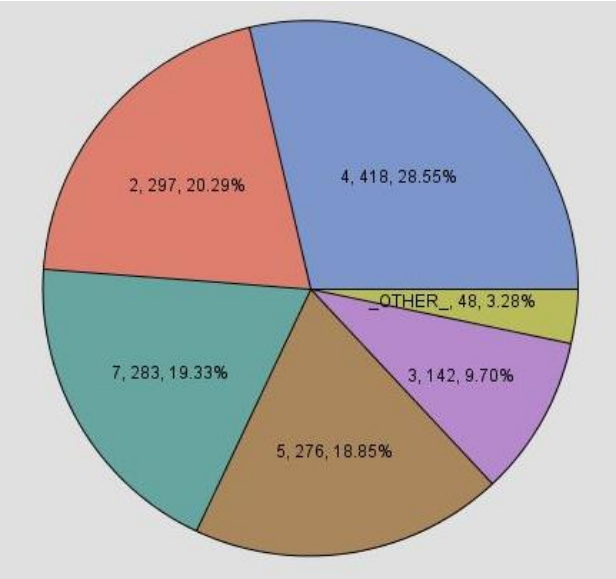
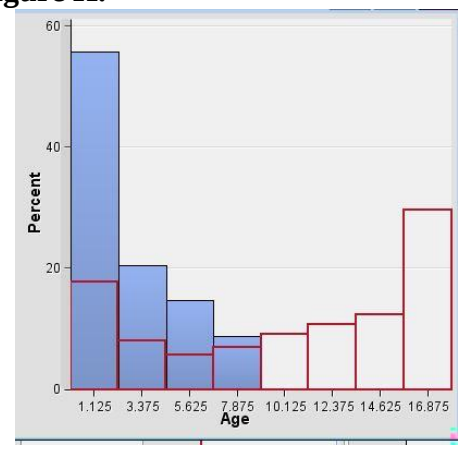


Figure H.



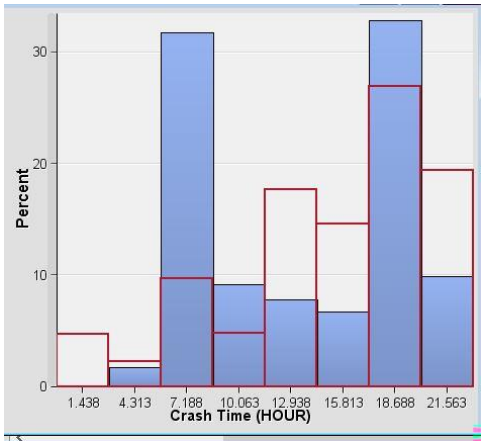
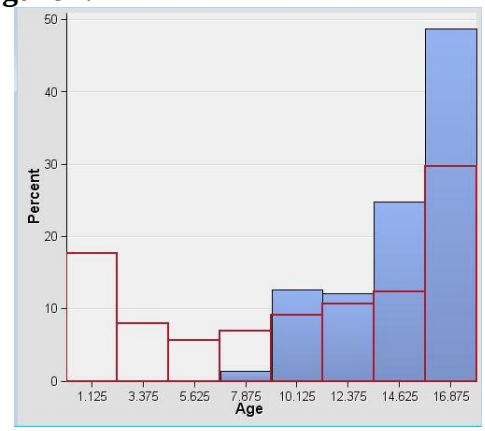


Figure I.



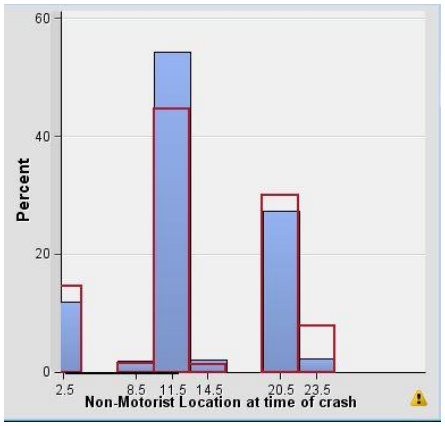


Figure J.

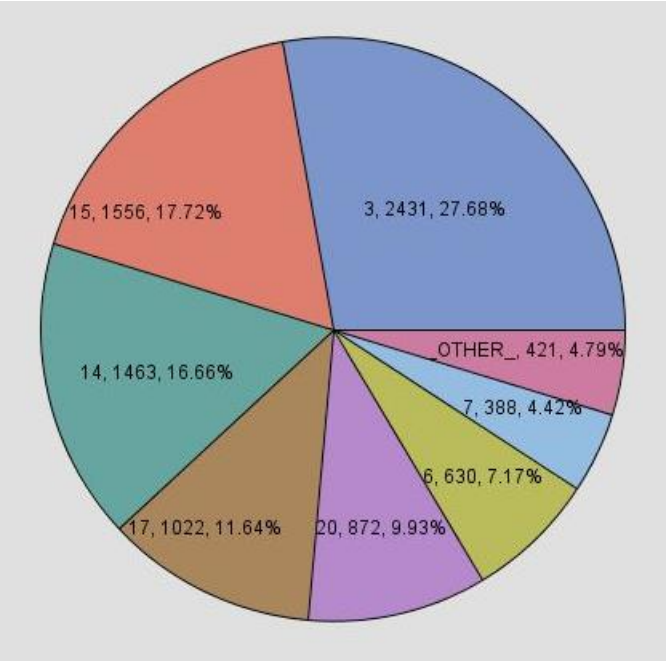


Figure K.

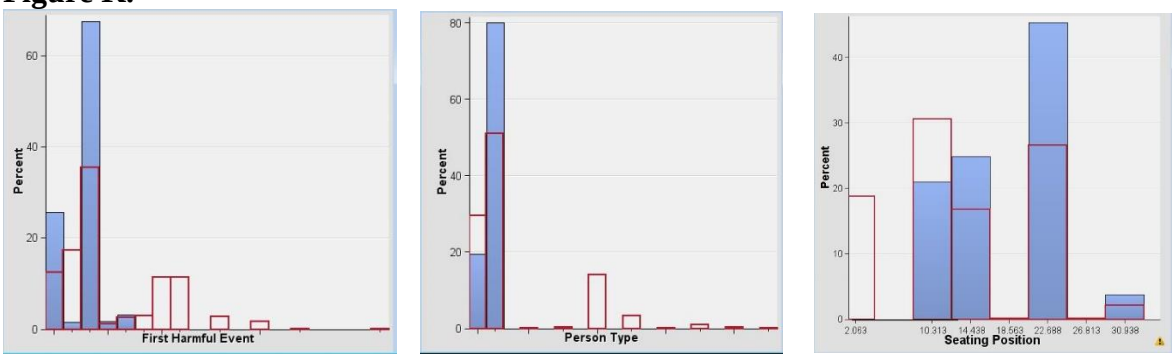


Figure L.

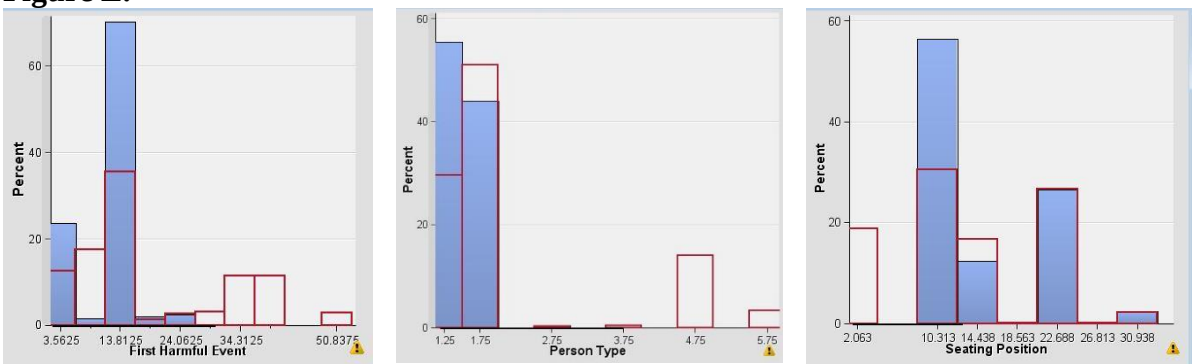


Figure M.

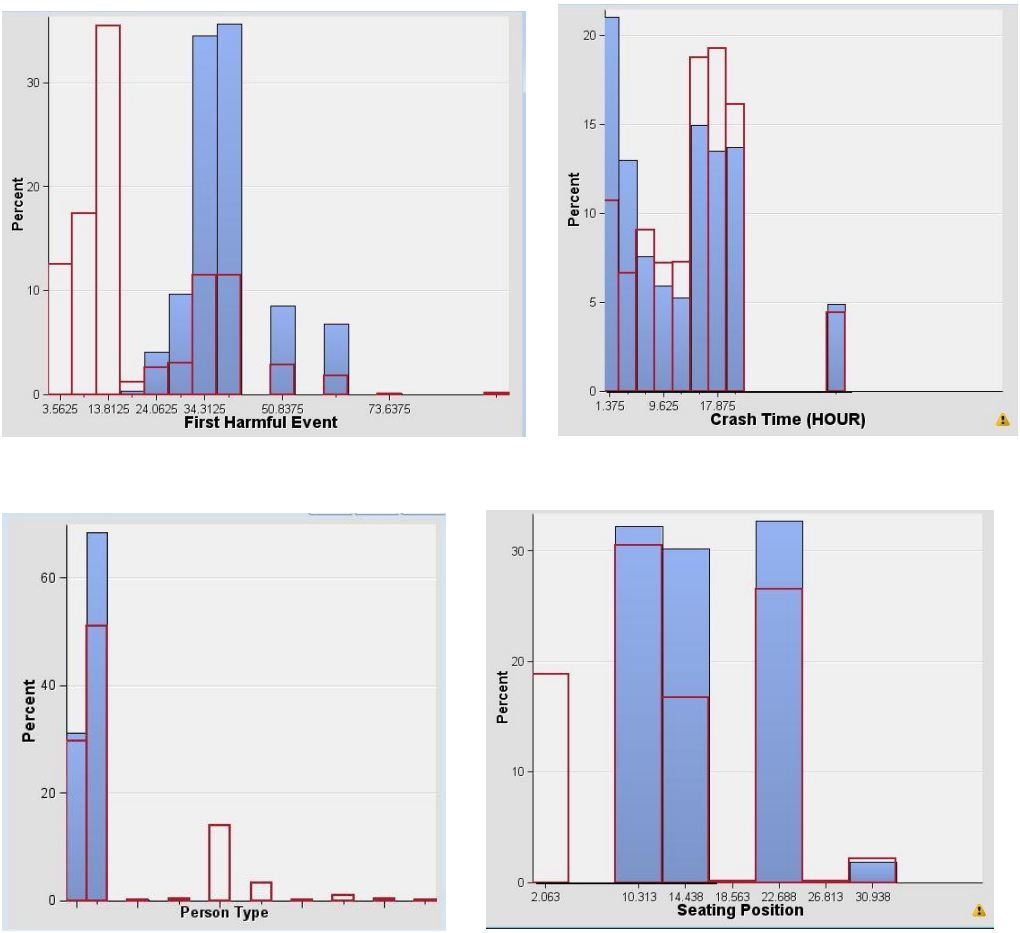


Figure N.

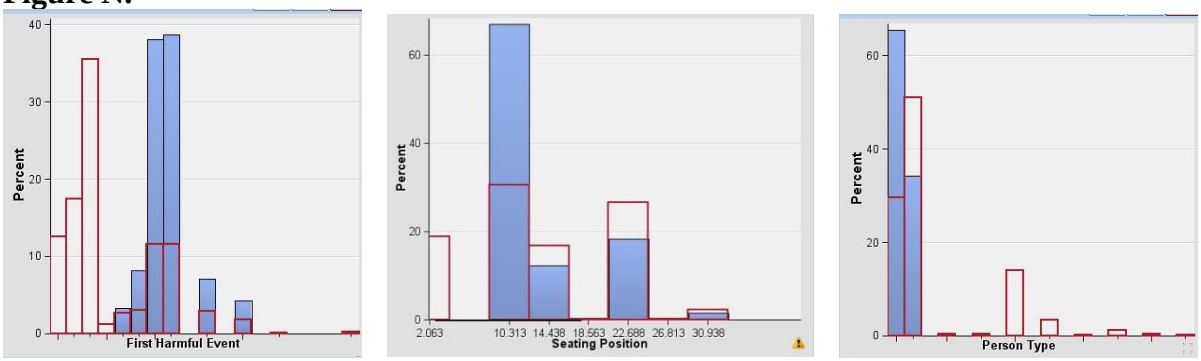


Figure O.

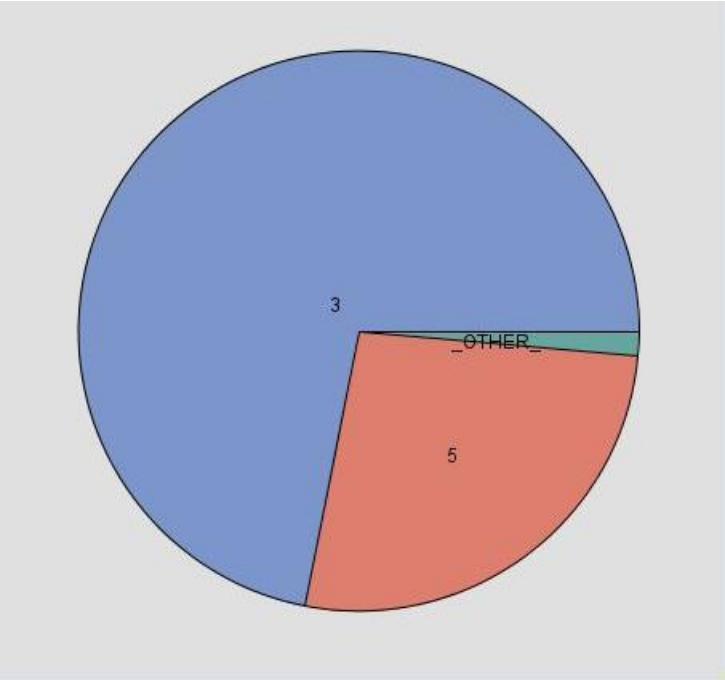
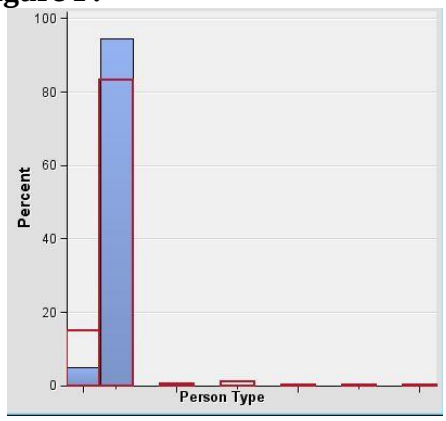


Figure P.



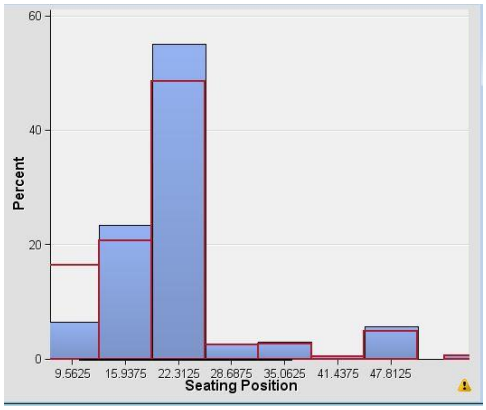
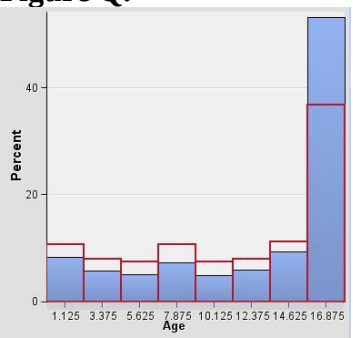
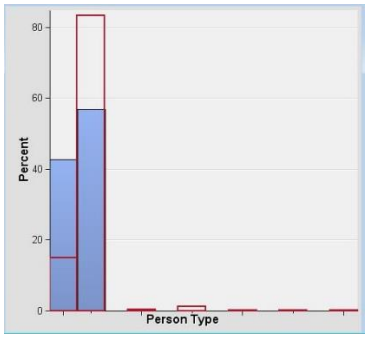


Figure Q.





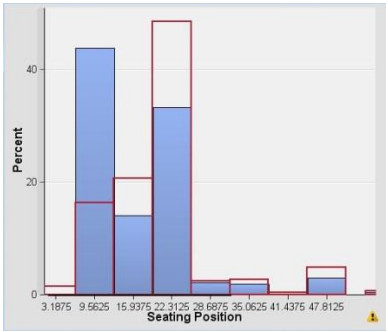


Figure R.

Deceased Non-Motorist Profiles

Educational Insight: Second youngest profile in the analysis, likely not alone when accident occurred...how can parental action improve in these situations, if any was taken at all?

Educational Insight: Old enough to be alone or with friends, hard to incentivize vests, other equipment

Survived Non-Motorist

Educational Insight: Mirror of profile from deceased non-motorists, similar insights just non-fatal injuries

Educational Insight: Parents taking their children outside before and after work… peak traffic hours; have to deal with distractions

Deceased In-Vehicle

CONNOR

- IN-VEHICLE
- COLLISION – MULTI-CAR
- BACK SEAT – PASSENGER

TARGET EDUCATOR:
TEENAGER: HEALTH/SAFETY CLASS

Educational Insight: Driver instinct to protect himself/herself when in collision event, risk to backseat passengers

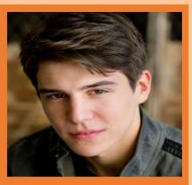
CAITLYN

- IN-VEHICLE
- COLLISION – MULTI-CAR
- FRONT SEAT – DRIVER

TARGET EDUCATOR:
TEENAGER: DRIVING SCHOOL

Educational Insight: Collision profile with more driver deaths, would support crossing the line/head on collisions

CHRIS



- IN-VEHICLE
- OFF ROAD OBSTACLE
- FRONT SEAT - PASSENGER
- LATE NIGHT

TARGET EDUCATOR:
TEENAGER: HEALTH/SAFETY CLASS

Educational Insight: Late night swerve crashes (earliest avg. crash time)... front seat passengers at most risk

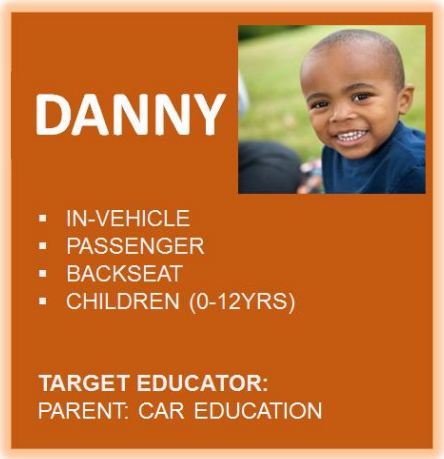
CARLY

- IN-VEHICLE
- OFF ROAD OBSTACLE
- FRONT SEAT – DRIVER
- ALONE

TARGET EDUCATOR:
TEENAGER: DRIVING SCHOOL

Educational Insight: Would support that the driver was alone (1.4% of those in the front seat of these crashes were passengers)

Survived In-Vehicle



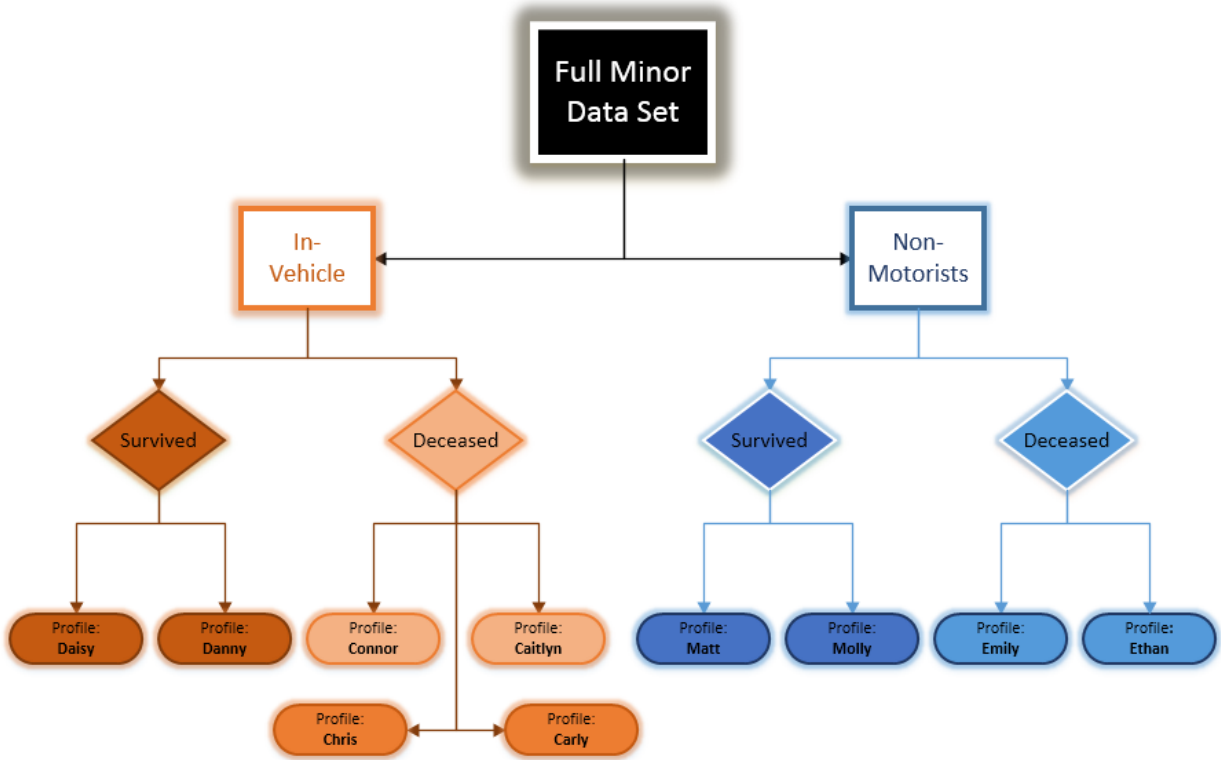
Educational Insight: Emphasizes the additional safety of the backseat for children, particularly when required by law

Educational Insight: Front seat drivers and passengers are severe minority compared to backseat passengers, reinforcement of enhanced back seat safety

Figure S.

FARS Data, 2011-2013, Age=<18			
	Non-Motorists	In-Vehicle	
Survived	MATT	DAISY	
	MOLLY	DANNY	
Deceased	EMILY	CONNOR	CHRIS
	ETHAN	CAITLYN	CARLY

Figure T.


Full Minor Data Set
In-Vehicle
Non-Motorists
Survived
Deceased
Survived
Deceased
Profile: Daisy
Profile: Danny
Profile: Connor
Profile: Caitlyn
Profile: Matt
Profile: Molly
Profile: Emily
Profile: Ethan
Profile: Chris
Profile: Carly