Preliminary Project Submission

Paul Jean-Marie

My Variables of Interest

Variable Name (in R)	Variable Type	Response/Explanatory	Variable Description
TRAV_SP	Numerical	Response	This element records the speed the vehicle was traveling prior to the occurrence of the crash as reported by the investigating officer in miles per hour (mph).
DOA	Dichotomous	Response	This element identifies if this person died at the scene of the crash or en route to a hospital or treatment facility. (Recoding – 7/8 – Died, 0 – Not Died)
AGE	Numerical	Explanatory	This element identifies the persons age, in years, with respect to the person's last birthday. (Note that 0 denotes Blank or Less than One year)
SPEEDREL	Dichotomous	Explanatory	This element identifies if the driver's speed was related to the crash as identified by law enforcement. (Recoding : 0 – Not Racing, 1 – Racing (2/3/4/5)
INJ_SEV	Categorical	Explanatory	This element describes the severity of the injury to this person in the crash.

Description of all recoding procedures:

TRAV_SP – The variable contains many none values denoted by '999'. We have converted them to null values and then imputed with the last observations. No recoding done for this variable. This is one response variable.

DOA – We are recoding it as 0 and 1. 0 will be for where death has not occurred. 7 (Death at spot) and 8 (Death en route to hospital) will be recoded to 1 (Death occurred).

AGE - The variable contains many none values denoted by '999'. We have converted them to null values and then imputed with the last observations. No recoding done for this variable. This is one response variable.

SPEEDREL – At first, the null values have been removed. We are recoding it as 0 and 1. 0 will be for 'No Racing'. All Racing cases will be recoded to 1.

INJ_SEV - The variable contains many none values denoted by '999'. We have converted them to null values and then imputed with the last observations. No recoding done for this variable. This is one response variable.

Research Questions

1. RQ 1 – Whether Age is any where related to the travel speed before accident?
   a. Rationale – If the age is less (young people), the people are more fun-loving and tend to drive in very high speed causing accidents.
2. RQ2 – Whether Racing or not related to the travel speed before accident?
   a. Rationale – If the racing happens, the speed is generally very high and hence it is supposed to cause accidents.
3. RQ 3 - Whether Severity of Injury is related to the travel speed before accident?
   a. Rationale – When accident happens in high speed, the severity of injury is more.
4. RQ 4 – Whether severity of Injury is related the death in the accident?
   a. Rationale – When the injury is severe, there is high chance of death to happen.
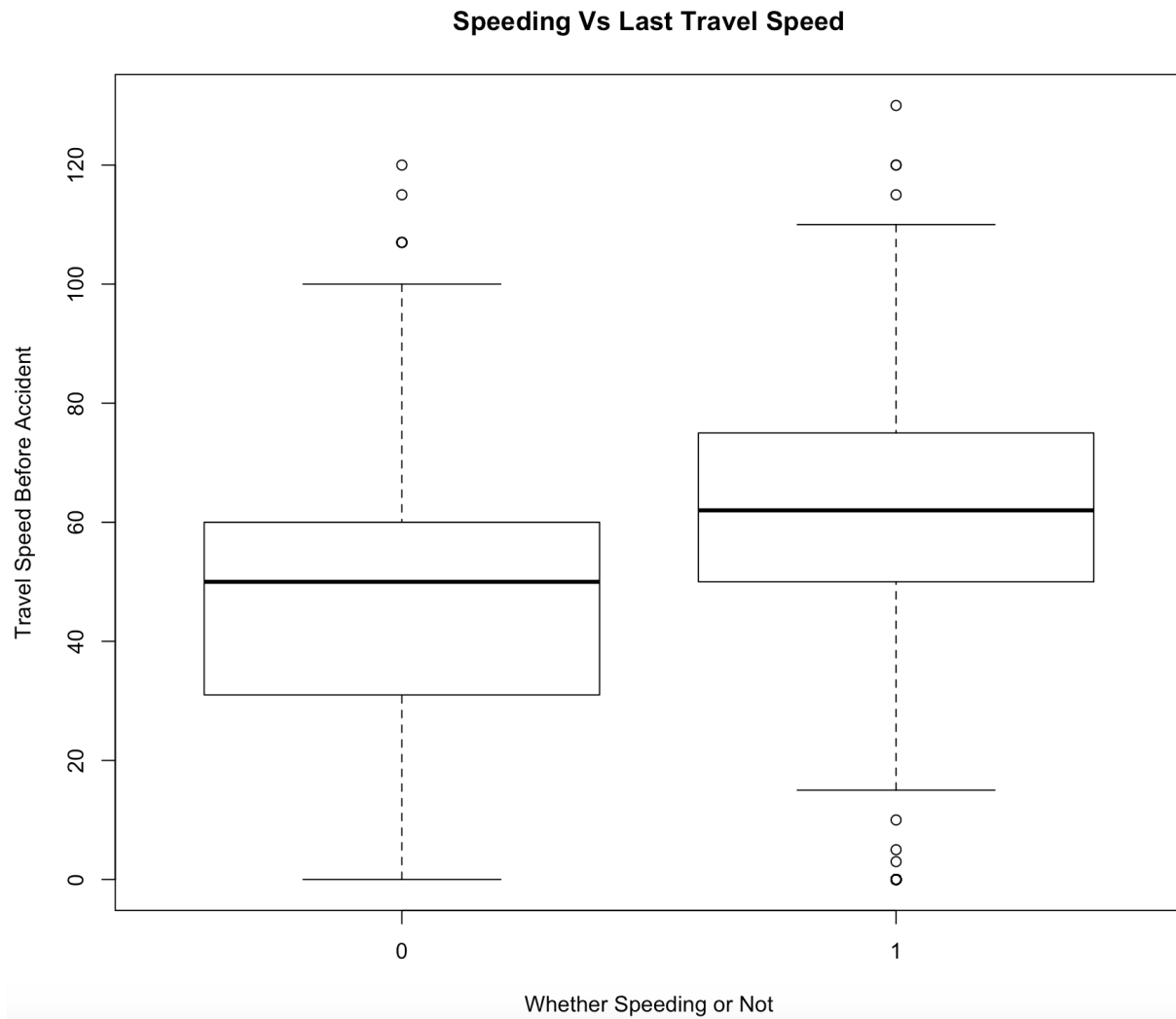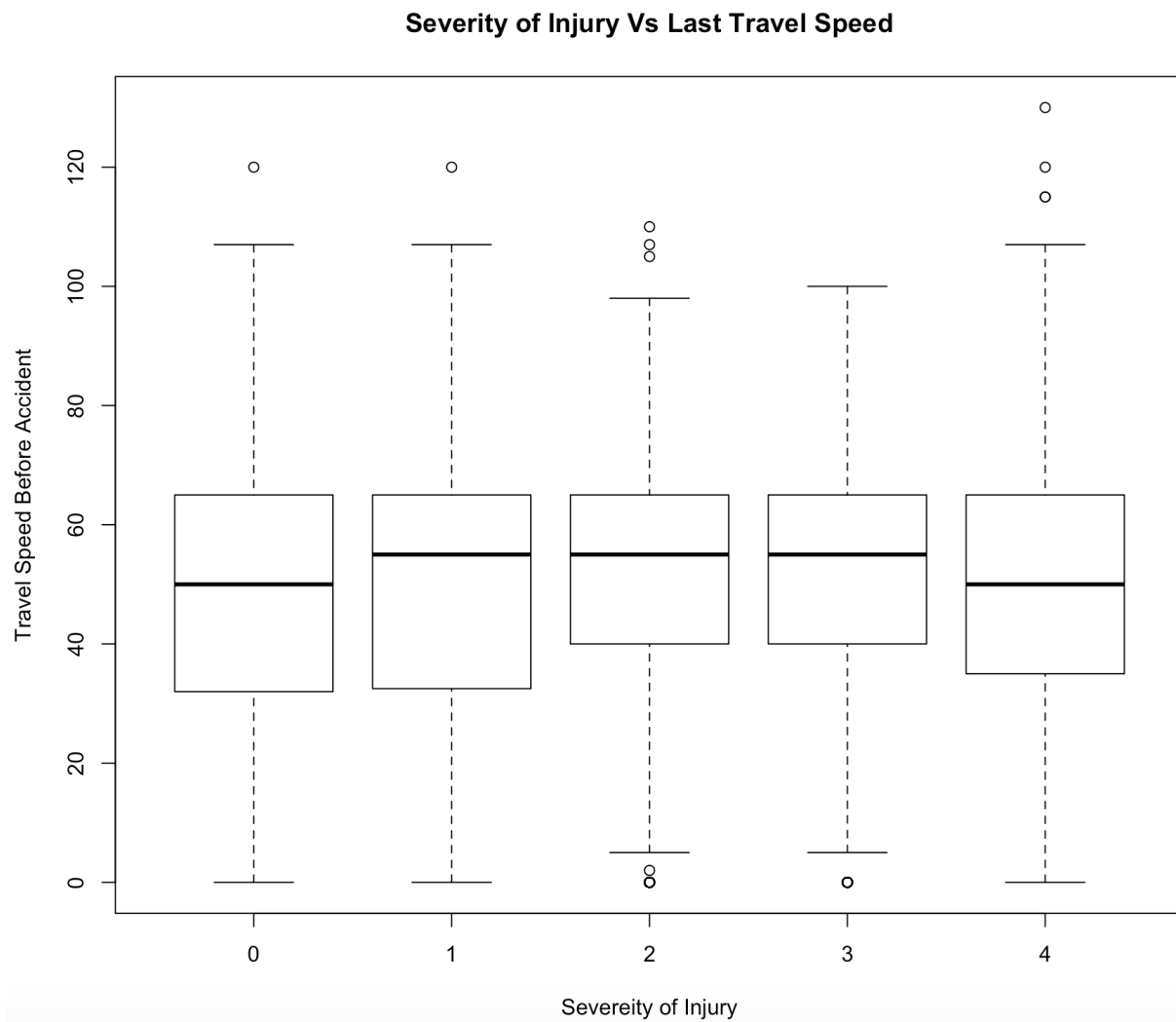
Plots

Figure 1: Age Vs Travel Speed
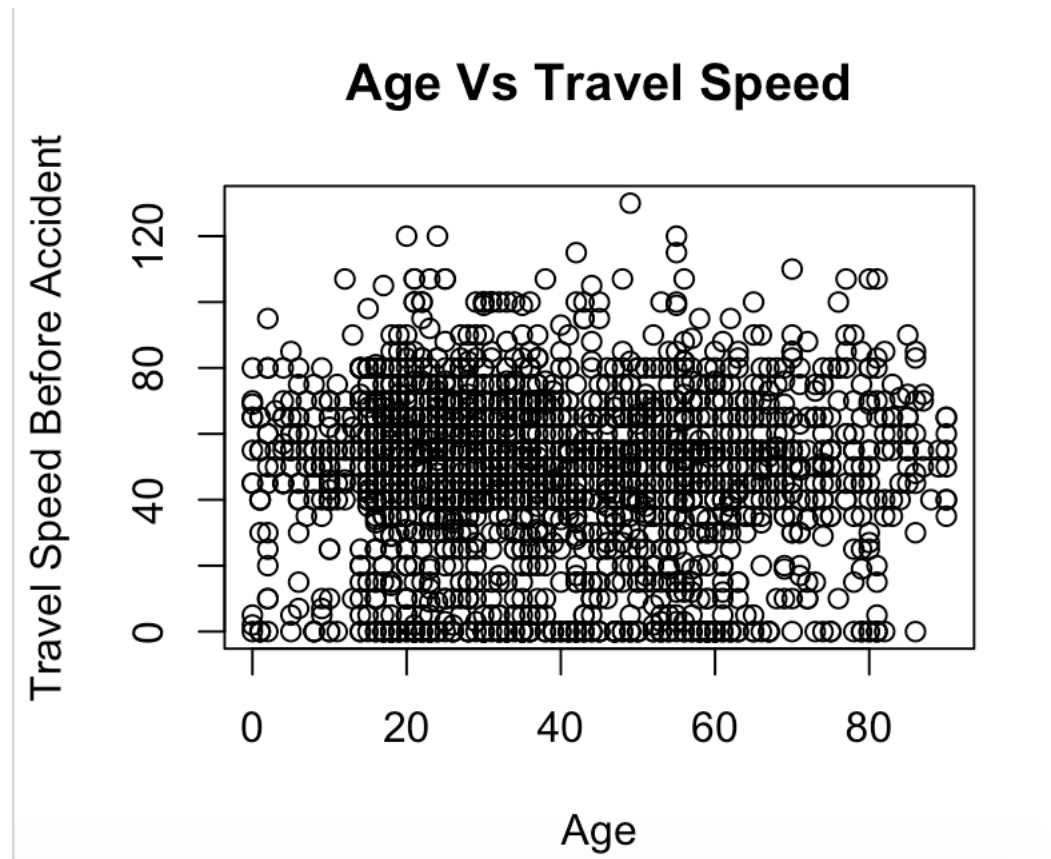
We are seeing that young people had travelled more in higher speeds before accidents.
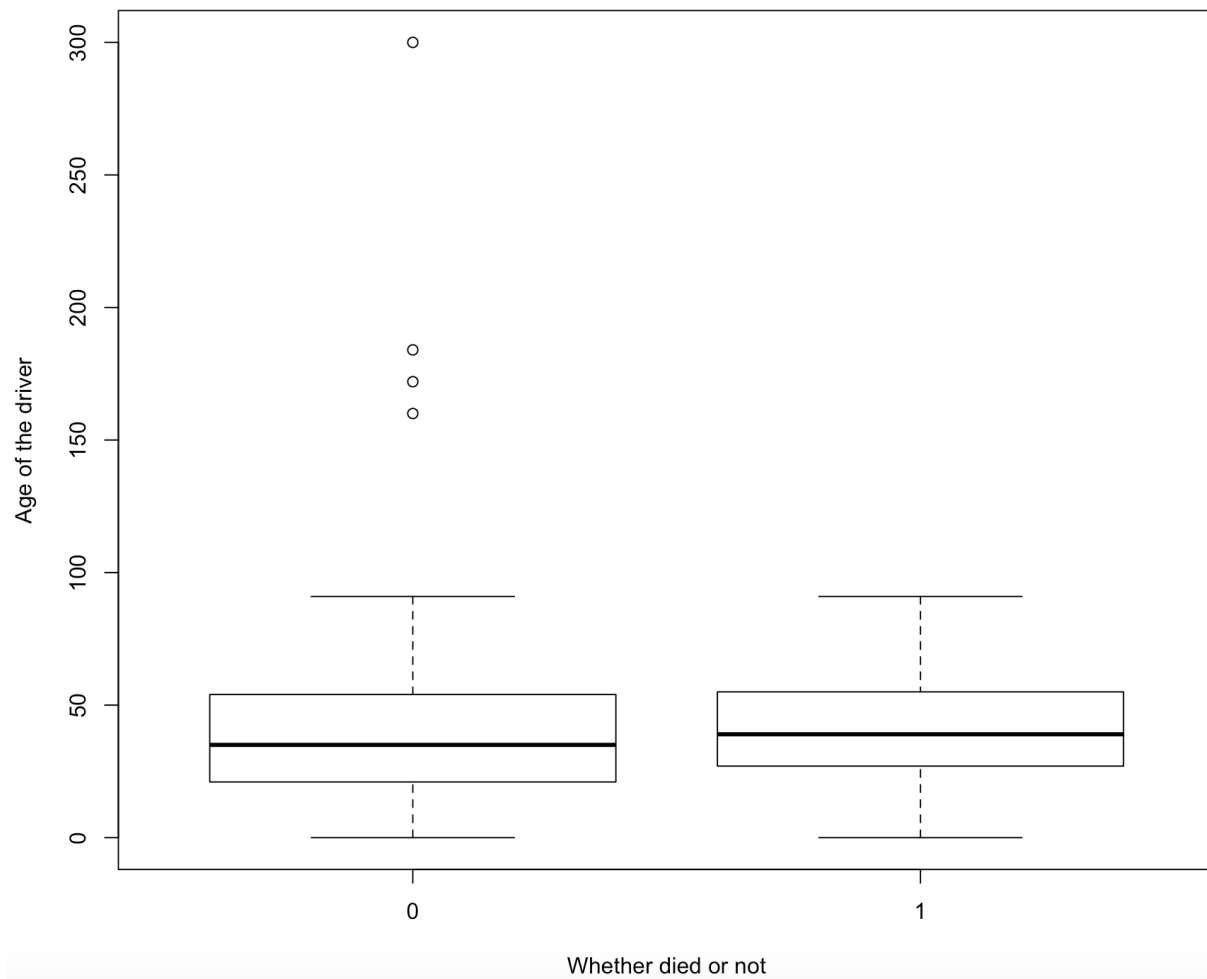
Figure 2: Racing Vs Travel Speed

Speeding Vs Last Travel Speed

When they are racing, the travel speed is more.

Figure 3: Severity of Injury Vs Travel Speed

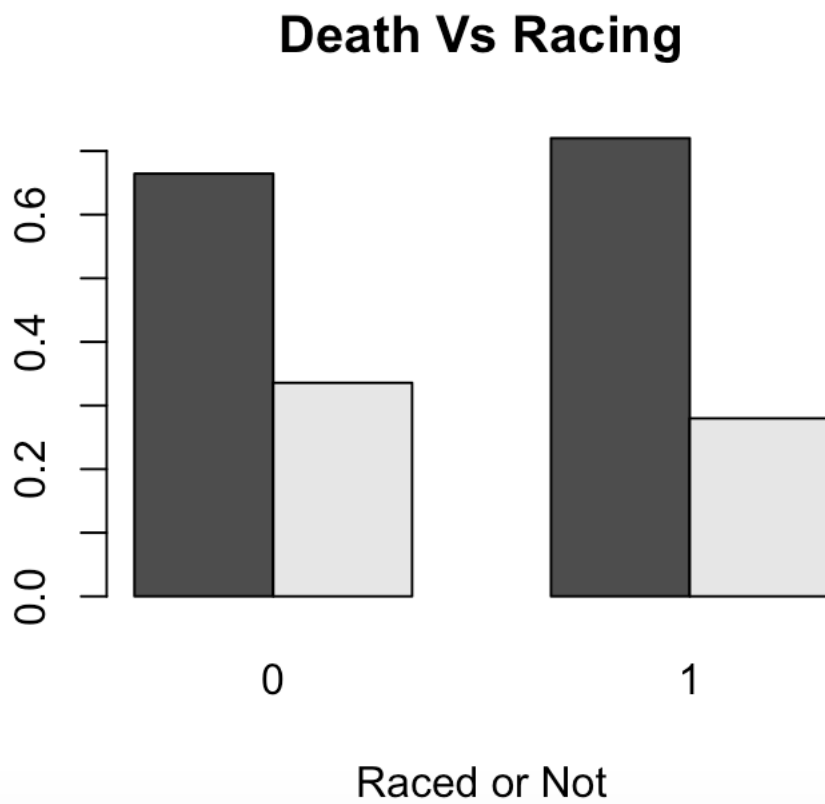**Severity of Injury Vs Last Travel Speed**

We are not seeing much difference here.

Figure 4: Whether Died or Not Vs Age of Driver

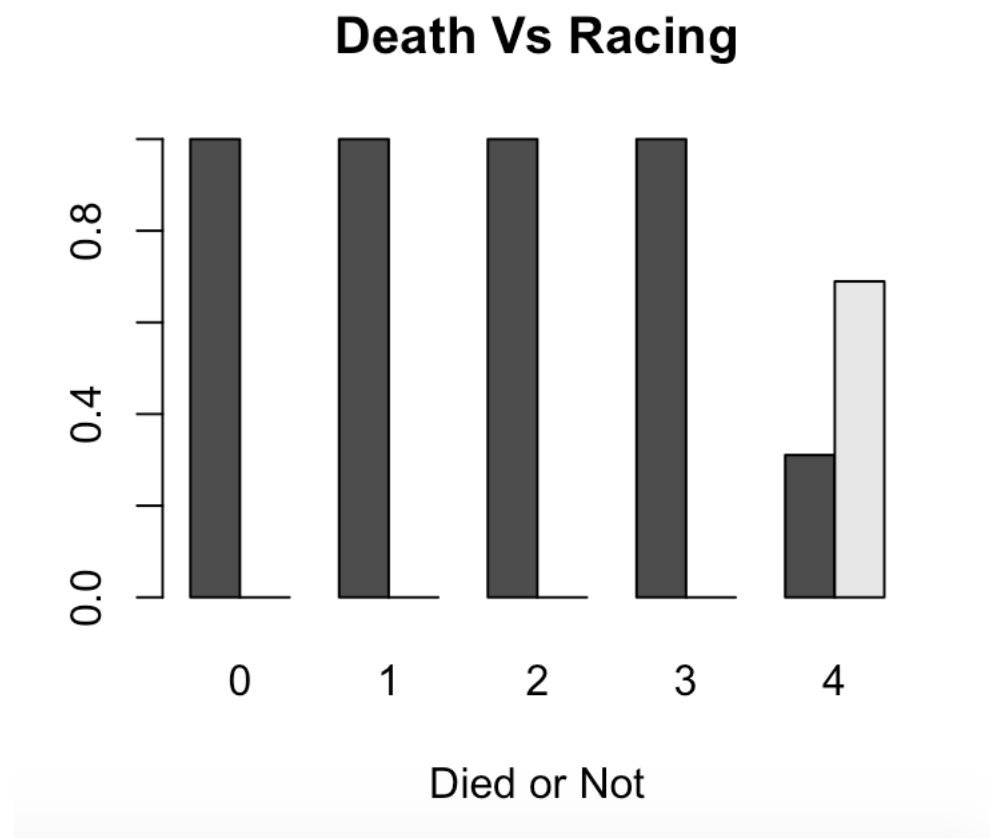We are not seeing much difference here.

Figure 5: Death Vs Racing

We are seeing deaths in case of racing is more.

Figure 6: Death Vs Severity Injury

Death Vs Racing

Died or Not

We are seeing those who are every much injured (4), they only died.

Works Cited

Appendix: R Code

Remember to submit your script as a separate .R file as well.

```
library(zoo)

accident <- read.csv('2018NHTSA1-SA-8258-1-4.csv')

#Imputation and Recoding
#TRAV_SP
accident$TRAV_SP <- ifelse(accident$TRAV_SP == 999, NA, accident$TRAV_SP)
accident$TRAV_SP <- na.locf(accident$TRAV_SP)

#DOA
accident$DOA <- ifelse(accident$DOA == 0, 0, 1)
accident$DOA <- factor(accident$DOA)

#AGE
accident$AGE <- ifelse(accident$AGE == 999, NA, accident$AGE)
accident$AGE <- na.locf(accident$AGE)
accident$AGE <- ifelse(accident$AGE >= 90, 90, accident$AGE)

#SPEEDREL
accident$SPEEDREL <- ifelse(accident$SPEEDREL == 999, NA, accident$SPEEDREL)
accident$SPEEDREL <- na.locf(accident$SPEEDREL)
accident$SPEEDREL <- ifelse(accident$SPEEDREL == 0, 0, 1)
accident$SPEEDREL <- factor(accident$SPEEDREL)

#INJ_SEV
accident$INJ_SEV <- ifelse(accident$INJ_SEV == 999, NA, accident$INJ_SEV)
accident$INJ_SEV <- na.locf(accident$INJ_SEV)
accident$INJ_SEV <- factor(accident$INJ_SEV)

#Plots
plot(accident$AGE, accident$TRAV_SP, ylab = 'Travel Speed Before Accident', xlab = 'Age', main =
'Age Vs Travel Speed' )

boxplot(accident$TRAV_SP ~ accident$SPEEDREL, xlab = 'Whether Speeding or Not', ylab = 'Travel
Speed Before Accident', main = 'Speeding Vs Last Travel Speed')

boxplot(accident$TRAV_SP ~ accident$INJ_SEV, xlab = 'Severeity of Injury', ylab = 'Travel Speed
Before Accident', main = 'Severity of Injury Vs Last Travel Speed')

boxplot(accident$AGE ~ accident$DOA, xlab = 'Whether died or not', ylab = 'Age of the driver')

table1 <- table(accident$DOA, accident$SPEEDREL)
proptable <- prop.table(table1, margin=2)

barplot(proptable, beside=TRUE, xlab = 'Raced or Not', main = 'Death Vs Racing' )

table1 <- table(accident$DOA, accident$INJ_SEV)
```

```
proptable <- prop.table(table1, margin=2)

barplot(proptable, beside=TRUE, xlab = 'Died or Not', main = 'Death Vs Racing' )
```