```
acci <- read.csv('C:\\Users\\hp\\Downloads\\Chrome\\accident.csv')
head(acci)
```

```
##   id  weight  dead airbag seatbelt frontal sex ageOFocc yearacc   abcat occRole
## 1  1  25.069 alive   none   belted       1   f       26    1997 unavail  driver
## 2  2  25.069 alive airbag   belted       1   f       72    1997  deploy  driver
## 3  3  32.379 alive   none     none       1   f       69    1997 unavail  driver
## 4  4 495.444 alive airbag   belted       1   f       53    1997  deploy  driver
## 5  5  25.069 alive   none   belted       1   f       32    1997 unavail  driver
## 6  6  25.069 alive   none   belted       1   f       22    1997 unavail  driver
##   deploy injSeverity  caseid
## 1      0           3 2:03:01
## 2      1           1 2:03:02
## 3      0           4 2:05:01
## 4      1           1 2:10:01
## 5      0           3 2:11:01
## 6      0           3 2:11:02
```

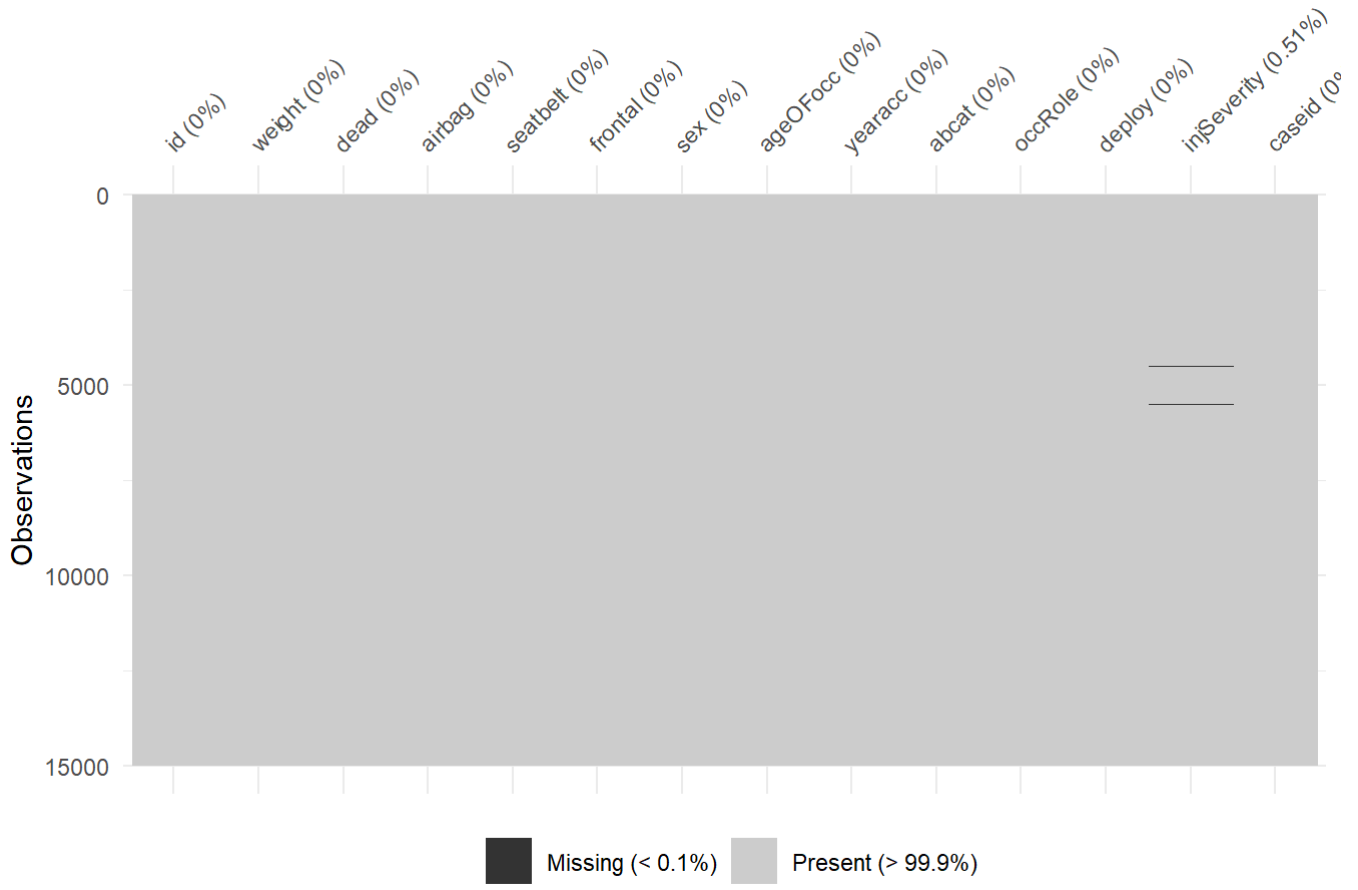
```
str(acci)
```

```
## 'data.frame':    14999 obs. of  14 variables:
##  $ id         : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ weight     : num  25.1 25.1 32.4 495.4 25.1 ...
##  $ dead       : chr  "alive" "alive" "alive" "alive" ...
##  $ airbag     : chr  "none" "airbag" "none" "airbag" ...
##  $ seatbelt   : chr  "belted" "belted" "none" "belted" ...
##  $ frontal    : int  1 1 1 1 1 1 1 0 1 ...
##  $ sex        : chr  "f" "f" "f" "f" ...
##  $ ageOFocc   : int  26 72 69 53 32 22 22 32 40 18 ...
##  $ yearacc    : int  1997 1997 1997 1997 1997 1997 1997 1997 1997 1997 ...
##  $ abcat      : chr  "unavail" "deploy" "unavail" "deploy" ...
##  $ occRole    : chr  "driver" "driver" "driver" "driver" ...
##  $ deploy     : int  0 1 0 1 0 0 0 0 0 0 ...
##  $ injSeverity: int  3 1 4 1 3 3 3 4 1 0 ...
##  $ caseid     : chr  "2:03:01" "2:03:02" "2:05:01" "2:10:01" ...
```

```
summary(acci)
```

```
##        id            weight             dead              airbag
##  Min.   :    1   Min.   :    0.00   Length:14999       Length:14999
##  1st Qu.: 3750   1st Qu.:   31.80   Class :character   Class :character
##  Median : 7500   Median :   82.98   Mode  :character   Mode  :character
##  Mean   : 7500   Mean   :  440.38
##  3rd Qu.:11250   3rd Qu.:  342.74
##  Max.   :14999   Max.   :57871.60
##
##    seatbelt            frontal           sex               ageOFocc
##  Length:14999       Min.   :0.0000   Length:14999       Min.   :16.00
##  Class :character   1st Qu.:0.0000   Class :character   1st Qu.:22.00
##  Mode  :character   Median :1.0000   Mode  :character   Median :33.00
##                     Mean   :0.6362                      Mean   :37.48
##                     3rd Qu.:1.0000                      3rd Qu.:48.00
##                     Max.   :1.0000                      Max.   :97.00
##
##     yearacc        abcat             occRole              deploy
##  Min.   :1997   Length:14999       Length:14999       Min.   :0.000
##  1st Qu.:1997   Class :character   Class :character   1st Qu.:0.000
##  Median :1998   Mode  :character   Mode  :character   Median :0.000
##  Mean   :1998                                         Mean   :0.294
##  3rd Qu.:1999                                         3rd Qu.:1.000
##  Max.   :2000                                         Max.   :1.000
##
##   injSeverity       caseid
##  Min.   :0.000   Length:14999
##  1st Qu.:1.000   Class :character
##  Median :2.000   Mode  :character
##  Mean   :1.746
##  3rd Qu.:3.000
##  Max.   :6.000
##  NA's   :76
```

```
library(naniar)
vis_miss(acci)
```




```
summary(is.na(acci))
```

```
##      id            weight          dead           airbag
##  Mode :logical   Mode :logical   Mode :logical   Mode :logical
##  FALSE:14999     FALSE:14999     FALSE:14999     FALSE:14999
##
##   seatbelt        frontal           sex           ageOFocc
##  Mode :logical   Mode :logical   Mode :logical   Mode :logical
##  FALSE:14999     FALSE:14999     FALSE:14999     FALSE:14999
##
##   yearacc          abcat          occRole          deploy
##  Mode :logical   Mode :logical   Mode :logical   Mode :logical
##  FALSE:14999     FALSE:14999     FALSE:14999     FALSE:14999
##
##  injSeverity       caseid
##  Mode :logical   Mode :logical
##  FALSE:14923     FALSE:14999
##  TRUE :76
```

```
head(acci$injSeverity)
```

```
## [1] 3 1 4 1 3 3
```

```
mode(acci$injSeverity)
```

```
## [1] "numeric"
```

```
mean(acci$injSeverity)
```

```
## [1] NA
```

```
acci$injSeverity
```

```
##    [1]  3  1  4  1  3  3  3  4  1  0  0  1  1  4  1  1  3  3  0  0  1  0  0  1
##   [25]  1  1  0  0  1  3  1  0  0  0  0  3  2  1  2  2  2  1  3  0  1  0  1  0
##   [49]  1  0  0  3  1  0  0  3  3  0  0  3  2  3  0  3  0  0  5  1  0  1  1  0
##   [73]  0  0  3  3  2  2  3  0  1  0  4  4  3  1  1  1  1  1  0  3  1  1  1  1
##   [97]  0  0  1  3  1  3  3  3  3  1  1  1  1  1  3  1  0  1  5  0  1  0  0  3
##  [121]  3  2  3  2  0  3  3  1  3  1  1  1  1  5  5  1  1  3  3  2  3  1  3  3
##  [145]  2  2  1  1  1  1  1  3  0  1  1  1  0  1  1  2  3  3  3  2  1  3  2  3
##  [169]  2  0  0  1  2  0  1  1  0  0  1  0  0  1  2  1  1  1  1  1  1  1  3  0
##  [193]  1  1  0  1  0  0  1  1  1  3  1  2  3  0  3  1  3  3  1  1  1  1  3  1
##  [217]  1  3  3  1  4  1  1  1  0  3  3  0  1  2  1  0  3  3  3  3  3  1  3  1
##  [241]  0  0  0  2  2  3  1  1  2  0  0  2  2  0  0  0  0  3  1  3  3  3  1  0
##  [265]  4  3  1  1  2  2  2  3  3  0  3  0  3  1  1  1  1  0  3  3  1  1  1  0
##  [289]  3  3  0  3  3  3  3  4  1  1  0  3  3  0  1  1  3  2  0  1  1  3  4  1
##  [313]  3  0  1  0  4  3  0  2  4  1  0  2  1  3  1  1  0  1  3  3  3  3  1  1
##  [337]  1  0  2  3  0  3  0  3  0  0  1  3  1  0  1  2  3  3  3  2  1  1  4  2
##  [361]  3  1  1  1  0  5  0  0  0  0  1  4  1  3  4  2  1  3  1  3  0  1  3  0
##  [385]  3  3  3  2  2  3  3  1  2  3  1  2  3  1  0  1  1  2  3  3  3  2  2  3
##  [409]  3  3  1  0  1  1  4  1  1  0  0  4  2  0  3  3  3  2  2  0  1  2  2  1
##  [433]  3  1  3  2  3  0  0  1  4  3  1  1  3  3  2  1  1  2  3  4  1  3  4  3
##  [457]  1  1  0  4  2  3  0  0  1  3  3  2  1  3  0  0  2  4  1  3  0  0  2  0
##  [481]  3  3  3  4  1  3  3  0  3  2  2  3  3  3  1  3  3  4  3  3  1  0  3  3
##  [505]  4  1  3  3  0  0  3  2  2  3  3  1  3  1  2  1  1  1  2  1  3  3  2  4
##  [529]  1  2  3  4  3  2  3  4  3  3  3  2  2  1  2  3  1  4  2  2  2  4  4  3
##  [553]  0  3  2  2  1  1  3  3  3  1  1  1  1  3  1  1  0  1  2  3  3  2  1  3
##  [577]  0  4  1  2  2  2  4  3  1  1  1  5  3  1  3  2  0  4  1  3  1  5  5  2
##  [601]  1  1  1  0  3  2  3  0  1  4  1  1  0  0  0  2  3  0  0  0  0  5  4  3
##  [625]  1  2  1  1  0  2  3  0  0  1  1  3  2  5 NA  0  0  1  1  0  1  0  0  0
##  [649]  3  1  5  0  0  3  5  3  3  3  3  2  3  3  1  3  3  1  2  2  0  0  2  1
##  [673]  1  2  4  3  0  2  0  0  1  0  3  1  1  0  3  0  0  1  2  3  3  2  1  0
##  [697]  0  1  1  1  1  0  5  3  1  2  1  3  1  2  0  2  0  0  0  0  4  3  3  0
##  [721]  0  0  5  2  1  2  5  2  3  1  2  3  4  1  0  1  1  2  0  0  2  3  3  2
##  [745]  2  3  2  3  1  1  1  5  5  0  3  1  3  0  1  0  0  1  1  1  0  1  1  0
##  [769]  0  0  1  3  1  5  3  3  3  3  0  2  1  0  1  1  1  0  3  3  0  0  2  3
##  [793]  1  3  1  1  2  1  0  0  0  0  3  3  0  0  0  2  1  0  1  0  1  3  0  5
##  [817]  2  3  2  2  3  0  0  1  3  2  4  1  3  1  2  3  3  0  3  3  3  4  2  3
##  [841]  2  3  3  3  3  3  3  3  3  2  2  3  3  0  3  0  3  3  1  1  3  1 NA  3
##  [865]  3  4  0  3  3  3  3  2  0  3  0  3  0  3  3  3  2  1  3  0  0  3  3  3
##  [889]  3  3  3  3  4  3  4  3  3  3  3  3  2  2  0  3  3  3  4  3  3  3  3  3
##  [913]  3  3  0  3  3  2  3  3  3  3  2  3  3  3  3  4  0  2  3  0  2  0  0  1
##  [937]  0  3  0  3  3  3  3  3  3  2  3  2  3  0  0  3  3  3  3  2  2  3  3  3
##  [961]  3  2  3  2  3  3  3  1  2  4  3  3  2  3  3  3  3  3  1  3  3  3  2  3
##  [985]  4  3  3  3  0  0  3  3  2  3  1  3  3  0  3  0  0  3  4  1  4  3  1  0
## [1009]  2  0  0  2  2  0  2  3  3  2  3  3  3  3  2  2  3  3  3  0  0  3  0  0
## [1033]  0  3  3  3  2  1  1  3  1  0  1  1  1  0  0  3  3  2  4  3  1  3  0  3
## [1057]  1  3  0  0  3  0  3  2  3  1  2  0  0  3  1  2  0  0  1  2  0  1  1  1
## [1081]  4  3  3  1  0  0  1  1  1  1  3  1  3  1  1  1  2  3  0  1  2  2  1  2
## [1105]  3  0  2  2  1  0  2  0  0  0  4  2  2  1  3  0  0  3  2  3  1  1  3  0
## [1129]  2  3  0  3  0  3  3  3  3  1  0  0  3  3  3  1  2  1  2  0  0  0  1  3
## [1153]  1  1  3  1  4  3  2  0  0  0  4  0  1  1  4  0  3  1  1  0  3  2  3  4
## [1177]  3  0  1  0  3  3  2  3  3  3  3  1  1  3  1  1  2  0  2  1  2  4  4  4
## [1201]  3  1  0  1  3  1  1 NA  2  0  3  0  0  0  0  0  3  3  3  1  3  0  3  4
## [1225]  3  0  0  3  0  2  2  1  0  0  0  0  0  1  4  0  1  3  0  3  4  0  3  3
## [1249]  3  2  0  0  0 NA  2  3  3  3  2  3  0  0  0  4  0  0  2  0  1  1  3  1
## [1273]  2  1  1  2  0  0  3  0  3  3  3  3  3  2  2  3  3  0  2  0  0  3  0  1
## [1297]  2  1  4  4  1  0  1  2  1  1  3  3  3  0  3  4  4  3  0  1  4  2  2  3
## [1321]  3  2  1  1  0  0  3  1  0  1  3  0  1  3  4  3  3  1  3  3  3 NA  2  2
## [1345]  2  3  1  2  0  0  1  0  0  6  3  1  1  0  1  2  1  0  1  0  0  0  2  1
```

```
##  [7124,] FALSE   FALSE  FALSE       FALSE  FALSE
##  [7125,] FALSE   FALSE  FALSE       FALSE  FALSE
##  [7126,] FALSE   FALSE  FALSE       FALSE  FALSE
##  [7127,] FALSE   FALSE  FALSE       FALSE  FALSE
##  [7128,] FALSE   FALSE  FALSE       FALSE  FALSE
##  [7129,] FALSE   FALSE  FALSE       FALSE  FALSE
##  [7130,] FALSE   FALSE  FALSE       FALSE  FALSE
##  [7131,] FALSE   FALSE  FALSE       FALSE  FALSE
##  [7132,] FALSE   FALSE  FALSE       FALSE  FALSE
##  [7133,] FALSE   FALSE  FALSE       FALSE  FALSE
##  [7134,] FALSE   FALSE  FALSE       FALSE  FALSE
##  [7135,] FALSE   FALSE  FALSE       FALSE  FALSE
##  [7136,] FALSE   FALSE  FALSE       FALSE  FALSE
##  [7137,] FALSE   FALSE  FALSE       FALSE  FALSE
##  [7138,] FALSE   FALSE  FALSE       FALSE  FALSE
##  [7139,] FALSE   FALSE  FALSE       FALSE  FALSE
##  [7140,] FALSE   FALSE  FALSE       FALSE  FALSE
##  [7141,] FALSE   FALSE  FALSE       FALSE  FALSE
##  [7142,] FALSE   FALSE  FALSE       FALSE  FALSE
##  [ reached getOption("max.print") -- omitted 7857 rows ]
```

```
sum(is.na(acci$injSeverity))
```

```
## [1] 76
```

```
acci$injSeverity[is.na(acci$injSeverity)] <- mean(acci$injSeverity,na.rm = T)
summary(acci)
```

```
##        id            weight            dead              airbag
##  Min.   :    1   Min.   :    0.00   Length:14999       Length:14999
##  1st Qu.: 3750   1st Qu.:   31.80   Class :character   Class :character
##  Median : 7500   Median :   82.98   Mode  :character   Mode  :character
##  Mean   : 7500   Mean   :  440.38
##  3rd Qu.:11250   3rd Qu.:  342.74
##  Max.   :14999   Max.   :57871.60
##    seatbelt            frontal           sex               ageOFocc
##  Length:14999       Min.   :0.0000   Length:14999       Min.   :16.00
##  Class :character   1st Qu.:0.0000   Class :character   1st Qu.:22.00
##  Mode  :character   Median :1.0000   Mode  :character   Median :33.00
##                     Mean   :0.6362                      Mean   :37.48
##                     3rd Qu.:1.0000                      3rd Qu.:48.00
##                     Max.   :1.0000                      Max.   :97.00
##     yearacc        abcat             occRole             deploy
##  Min.   :1997   Length:14999       Length:14999       Min.   :0.000
##  1st Qu.:1997   Class :character   Class :character   1st Qu.:0.000
##  Median :1998   Mode  :character   Mode  :character   Median :0.000
##  Mean   :1998                                         Mean   :0.294
##  3rd Qu.:1999                                         3rd Qu.:1.000
##  Max.   :2000                                         Max.   :1.000
##   injSeverity       caseid
##  Min.   :0.000   Length:14999
##  1st Qu.:1.000   Class :character
##  Median :2.000   Mode  :character
##  Mean   :1.746
##  3rd Qu.:3.000
##  Max.   :6.000
```

```
sum(is.na(acci$injSeverity))
```

```
## [1] 0
```

```
attach(acci)

library(ggplot2)

# store in factor

acci$dead = as.factor(acci$dead)
levels(acci$dead)
```

```
## [1] "alive" "dead"
```

```
#create new category variable

dead_alive <- ifelse(acci$dead == 'dead',0,1)
dead_alive
```

```
##    [1] 1 1 1 1 1 1 1 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##   [37] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##   [73] 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [109] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [145] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [181] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [217] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [253] 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [289] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 0 1 1 1 0 1 1 1
##  [325] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1
##  [361] 1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [397] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1
##  [433] 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 0 1 1 0 1 1 1 1 0 1 1 1 1 1 1 1 1
##  [469] 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 0 1 0 1 1 1 1 1 0 1 1 1 1 1 0 1 1 1 1 1
##  [505] 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 0 1 1 1 0 1 1 1
##  [541] 1 1 1 1 1 0 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [577] 1 0 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1
##  [613] 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [649] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1
##  [685] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1
##  [721] 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [757] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [793] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1
##  [829] 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [865] 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 1 1 1 1 1
##  [901] 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1
##  [937] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1
##  [973] 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 1 1 1
## [1009] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [1045] 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [1081] 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1
## [1117] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [1153] 1 1 1 1 0 1 1 1 1 1 0 1 1 1 0 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1
## [1189] 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0
## [1225] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [1261] 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [1297] 1 1 0 0 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [1333] 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [1369] 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1
## [1405] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 0 1 1 0 0 1 0 1 1 1 1
## [1441] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1
## [1477] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1
## [1513] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 0 1 1
## [1549] 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1
## [1585] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [1621] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [1657] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 0 1 1 0 1 1 1
## [1693] 1 1 0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [1729] 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [1765] 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [1801] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1
## [1837] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1
## [1873] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0
## [1909] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1
## [1945] 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1
## [1981] 1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [2017] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

```
## [14581] 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [14617] 1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [14653] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1
## [14689] 1 1 1 1 1 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1
## [14725] 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1
## [14761] 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1
## [14797] 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [14833] 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1
## [14869] 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1
## [14905] 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [14941] 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [14977] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

```
acci <- cbind(acci,dead_alive)
acci
```

```
##        id    weight  dead airbag seatbelt frontal sex ageOFocc yearacc    abcat
## 1       1    25.069 alive   none   belted       1   f       26    1997  unavail
## 2       2    25.069 alive airbag   belted       1   f       72    1997   deploy
## 3       3    32.379 alive   none     none       1   f       69    1997  unavail
## 4       4   495.444 alive airbag   belted       1   f       53    1997   deploy
## 5       5    25.069 alive   none   belted       1   f       32    1997  unavail
## 6       6    25.069 alive   none   belted       1   f       22    1997  unavail
## 7       7    27.078 alive   none   belted       1   m       22    1997  unavail
## 8       8    27.078  dead   none     none       1   m       32    1997  unavail
## 9       9   812.869 alive   none   belted       0   m       40    1997  unavail
## 10     10   812.869 alive   none   belted       1   f       18    1997  unavail
## 11     11   812.869 alive   none   belted       1   m       21    1997  unavail
## 12     12   299.459 alive   none   belted       0   f       69    1997  unavail
## 13     13   923.996 alive airbag   belted       1   m       67    1997   deploy
## 14     14    89.627  dead airbag   belted       0   f       54    1997 nodeploy
## 15     15    29.081 alive   none   belted       0   f       37    1997  unavail
## 16     16    29.081 alive   none   belted       1   m       71    1997  unavail
## 17     17    29.081 alive   none   belted       1   f       63    1997  unavail
## 18     18    58.191 alive   none     none       0   f       23    1997  unavail
## 19     19  1156.439 alive airbag   belted       0   m       33    1997 nodeploy
## 20     20  1156.439 alive airbag   belted       0   m       17    1997 nodeploy
## 21     21    86.200 alive airbag     none       1   f       20    1997   deploy
## 22     22    86.200 alive   none   belted       0   f       22    1997  unavail
## 23     23   397.171 alive   none     none       0   m       45    1997  unavail
## 24     24   397.171 alive   none     none       0   m       36    1997  unavail
## 25     25    37.099 alive airbag   belted       1   f       23    1997   deploy
## 26     26    79.205 alive   none     none       1   f       19    1997  unavail
## 27     27   244.390 alive   none   belted       0   m       26    1997  unavail
## 28     28   244.390 alive   none   belted       1   f       41    1997  unavail
## 29     29    26.233 alive   none   belted       1   f       18    1997  unavail
## 30     30    26.233 alive   none     none       1   m       30    1997  unavail
## 31     31   241.561 alive airbag   belted       1   f       26    1997   deploy
## 32     32   745.351 alive   none   belted       1   m       16    1997  unavail
## 33     33   745.351 alive   none   belted       1   m       17    1997  unavail
## 34     34   745.351 alive airbag   belted       0   m       20    1997 nodeploy
## 35     35   745.351 alive airbag   belted       0   f       17    1997 nodeploy
## 36     36    30.148 alive   none     none       1   m       32    1997  unavail
## 37     37    30.148 alive airbag     none       1   f       19    1997   deploy
## 38     38    30.148 alive airbag   belted       0   f       52    1997 nodeploy
## 39     39   199.316 alive   none   belted       1   f       64    1997  unavail
## 40     40   199.316 alive   none   belted       1   f       19    1997  unavail
## 41     41   199.316 alive   none   belted       1   m       36    1997  unavail
## 42     42   199.316 alive airbag   belted       0   m       23    1997 nodeploy
## 43     43    25.069 alive   none     none       1   m       76    1997  unavail
## 44     44   206.309 alive airbag   belted       1   m       32    1997   deploy
## 45     45   206.309 alive   none   belted       0   m       21    1997  unavail
## 46     46   206.309 alive   none   belted       0   m       25    1997  unavail
## 47     47   440.462 alive airbag   belted       0   f       37    1997 nodeploy
## 48     48   440.462 alive airbag     none       0   f       16    1997 nodeploy
## 49     49   105.152 alive   none   belted       0   f       24    1997  unavail
## 50     50   105.152 alive   none   belted       1   m       39    1997  unavail
## 51     51   324.453 alive   none   belted       1   m       18    1997  unavail
## 52     52    98.523 alive   none   belted       1   f       64    1997  unavail
## 53     53   435.568 alive   none   belted       1   m       22    1997  unavail
## 54     54  1343.968 alive   none   belted       1   m       27    1997  unavail
## 55     55    29.036 alive   none   belted       1   m       18    1997  unavail
## 56     56    29.036 alive   none   belted       0   f       20    1997  unavail
```

```
## 6614  driver      0    0.000000     45:90:1          1
## 6615  driver      0    3.000000     45:90:2          1
## 6616  driver      0    0.000000     45:92:1          1
## 6617  driver      0    0.000000     45:92:2          1
## 6618    pass      0    0.000000     45:92:2          1
## 6619  driver      0    2.000000     45:93:1          1
## 6620    pass      0    4.000000     45:93:1          0
## 6621  driver      0    3.000000     45:94:1          1
## 6622  driver      0    1.000000     45:96:1          1
## 6623  driver      0    1.000000     45:97:1          1
## 6624  driver      1    1.000000     45:98:1          1
## 6625  driver      0    3.000000     45:98:2          1
## 6626  driver      1    3.000000     45:99:1          1
## 6627  driver      0    3.000000     45:99:2          1
## 6628  driver      0    3.000000    45:103:1          1
## 6629    pass      0    3.000000    45:103:1          1
## 6630  driver      0    2.000000    45:103:2          1
## 6631  driver      0    0.000000    45:103:3          1
## 6632  driver      0    3.000000    45:104:1          1
## 6633    pass      0    0.000000    45:104:1          1
## 6634  driver      0    0.000000    45:104:2          1
## 6635  driver      0    3.000000    45:106:1          1
## 6636    pass      0    1.000000    45:106:1          1
## 6637  driver      0    0.000000    45:107:2          1
## 6638    pass      0    2.000000    45:107:2          1
## 6639  driver      1    2.000000    45:108:1          1
## 6640  driver      0    2.000000    45:108:2          1
## 6641    pass      0    2.000000    45:108:2          1
## 6642  driver      0    3.000000    45:109:2          1
## 6643  driver      0    2.000000    45:110:1          1
## 6644  driver      0    3.000000    45:110:2          1
## 6645    pass      0    3.000000    45:110:2          1
## 6646  driver      1    3.000000    45:111:1          1
## 6647    pass      1    3.000000    45:111:1          1
## 6648  driver      0    3.000000    45:111:2          1
## 6649  driver      0    4.000000    45:114:1          0
## 6650    pass      0    3.000000    45:114:1          1
## 6651  driver      1    1.000000    45:114:2          1
## 6652  driver      0    3.000000    45:116:1          1
## 6653  driver      1    2.000000    45:116:2          1
## 6654  driver      0    3.000000    45:118:1          1
## 6655    pass      0    3.000000    45:118:1          1
## 6656  driver      0    2.000000    45:118:2          1
## 6657  driver      1    1.000000    45:119:1          1
## 6658  driver      0    3.000000    45:119:2          1
## 6659  driver      1    0.000000    45:120:1          1
## 6660  driver      0    0.000000    45:120:2          1
## 6661  driver      1    2.000000    45:121:1          1
## 6662    pass      1    3.000000    45:121:1          1
## 6663  driver      0    3.000000    45:122:1          1
## 6664  driver      0    3.000000    45:122:2          1
## 6665  driver      0    3.000000    45:123:1          1
## 6666  driver      0    2.000000    45:123:2          1
##  [ reached 'max' / getOption("max.print") -- omitted 8333 rows ]
```

```
str(acci)
```

```
## 'data.frame':    14999 obs. of  15 variables:
##  $ id         : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ weight     : num  25.1 25.1 32.4 495.4 25.1 ...
##  $ dead       : Factor w/ 2 levels "alive","dead": 1 1 1 1 1 1 1 2 1 1 ...
##  $ airbag     : chr  "none" "airbag" "none" "airbag" ...
##  $ seatbelt   : chr  "belted" "belted" "none" "belted" ...
##  $ frontal    : int  1 1 1 1 1 1 1 1 0 1 ...
##  $ sex        : chr  "f" "f" "f" "f" ...
##  $ ageOFocc   : int  26 72 69 53 32 22 22 32 40 18 ...
##  $ yearacc    : int  1997 1997 1997 1997 1997 1997 1997 1997 1997 1997 ...
##  $ abcat      : chr  "unavail" "deploy" "unavail" "deploy" ...
##  $ occRole    : chr  "driver" "driver" "driver" "driver" ...
##  $ deploy     : int  0 1 0 1 0 0 0 0 0 0 ...
##  $ injSeverity: num  3 1 4 1 3 3 3 4 1 0 ...
##  $ caseid     : chr  "2:03:01" "2:03:02" "2:05:01" "2:10:01" ...
##  $ dead_alive : num  1 1 1 1 1 1 1 0 1 1 ...
```

```
acci_cols <- sapply(acci,is.numeric)
acci <- acci[,acci_cols]
head(acci)
```

```
##   id  weight frontal ageOFocc yearacc deploy injSeverity dead_alive
## 1  1  25.069       1       26    1997      0           3          1
## 2  2  25.069       1       72    1997      1           1          1
## 3  3  32.379       1       69    1997      0           4          1
## 4  4 495.444       1       53    1997      1           1          1
## 5  5  25.069       1       32    1997      0           3          1
## 6  6  25.069       1       22    1997      0           3          1
```

```
cor(acci)
```

```
##                      id       weight      frontal     ageOFocc      yearacc
## id           1.000000000  0.009455983  0.003460235  0.009996999  0.960219362
## weight       0.009455983  1.000000000 -0.014071898 -0.026647965  0.002458269
## frontal      0.003460235 -0.014071898  1.000000000 -0.038994758 -0.005061151
## ageOFocc     0.009996999 -0.026647965 -0.038994758  1.000000000  0.016792323
## yearacc      0.960219362  0.002458269 -0.005061151  0.016792323  1.000000000
## deploy       0.051074730 -0.051855705  0.200286430  0.018657016  0.059793522
## injSeverity -0.032962207 -0.195518745 -0.024958986  0.086942131 -0.043907552
## dead_alive   0.016927290  0.057623687  0.065044486 -0.092837318  0.018174919
##                  deploy injSeverity  dead_alive
## id           0.05107473 -0.03296221  0.01692729
## weight      -0.05185570 -0.19551874  0.05762369
## frontal      0.20028643 -0.02495899  0.06504449
## ageOFocc     0.01865702  0.08694213 -0.09283732
## yearacc      0.05979352 -0.04390755  0.01817492
## deploy       1.00000000  0.05773224  0.01976222
## injSeverity  0.05773224  1.00000000 -0.38062763
## dead_alive   0.01976222 -0.38062763  1.00000000
```

```
library(caTools)
set.seed(101)
split <- sample.split(acci$dead_alive,SplitRatio = 0.7)

final.train <- subset(acci,split = TRUE)
final.test <- subset(acci,split = FALSE)

# train the data
str(final.train)
```

```
## 'data.frame':    14999 obs. of  8 variables:
##  $ id         : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ weight     : num  25.1 25.1 32.4 495.4 25.1 ...
##  $ frontal    : int  1 1 1 1 1 1 1 1 0 1 ...
##  $ ageOFocc   : int  26 72 69 53 32 22 22 32 40 18 ...
##  $ yearacc    : int  1997 1997 1997 1997 1997 1997 1997 1997 1997 1997 ...
##  $ deploy     : int  0 1 0 1 0 0 0 0 0 0 ...
##  $ injSeverity: num  3 1 4 1 3 3 3 4 1 0 ...
##  $ dead_alive : num  1 1 1 1 1 1 1 0 1 1 ...
```

```
str(final.test)
```

```
## 'data.frame':    14999 obs. of  8 variables:
##  $ id         : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ weight     : num  25.1 25.1 32.4 495.4 25.1 ...
##  $ frontal    : int  1 1 1 1 1 1 1 1 0 1 ...
##  $ ageOFocc   : int  26 72 69 53 32 22 22 32 40 18 ...
##  $ yearacc    : int  1997 1997 1997 1997 1997 1997 1997 1997 1997 1997 ...
##  $ deploy     : int  0 1 0 1 0 0 0 0 0 0 ...
##  $ injSeverity: num  3 1 4 1 3 3 3 4 1 0 ...
##  $ dead_alive : num  1 1 1 1 1 1 1 0 1 1 ...
```

```
#train the model

final.model <- glm(formula = dead_alive ~ weight +ageOFocc + injSeverity , family = binomial
(link='logit'),data = final.train)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(final.model)
```

```
##
## Call:
## glm(formula = dead_alive ~ weight + ageOFocc + injSeverity, family = binomial(link = "logi
t"),
##     data = final.train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -6.2247   0.0000   0.0012   0.1611   5.0560
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 19.708626   0.532181  37.034  < 2e-16 ***
## weight       0.011741   0.001045  11.238  < 2e-16 ***
## ageOFocc    -0.017156   0.003587  -4.782 1.73e-06 ***
## injSeverity -5.269680   0.143793 -36.648  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5759.5  on 14998  degrees of freedom
## Residual deviance: 1608.9  on 14995  degrees of freedom
## AIC: 1616.9
##
## Number of Fisher Scoring iterations: 11
```

```
#create predict for test data

fitted.prob <- predict(final.model,newdata = final.test,type = 'response')

fitted.result = ifelse(fitted.prob >0.5,1,0)

#predict value in probability of 0 and 1

#accuracy

misclassicerror = mean(fitted.result<-final.test$dead_alive,na.rm = T)

print(paste(1-misclassicerror))
```

```
## [1] "0.0478031868791253"
```

```
print(paste('Accuracy is' , misclassicerror))
```

```
## [1] "Accuracy is 0.952196813120875"
```

```
#Accuracy is 0.9521

table(final.test$dead_alive,fitted.prob>0.5)
```

```
##
##     FALSE  TRUE
##   0   650    67
##   1    72 14210
```