

Assignment 2

Michael LaVallee

9/18/2019

1. Download the c2015 dataset to your computer. Use function `getwd()` to check the current working directory. Use `setwd()` to change the current directory to the c2015 file

```
getwd()
```

```
## [1] "C:/Users/student/Documents/Senior Year/Fall Semester/R Analytics/Assignment 2"
```

```
#my directory is already set i do not need to change it
```

2. We need to install a package to read the xlsx file. (Let's not change the xlsx to csv here) There are a few packages for this. I recommend to use the r

```
library(readxl)
```

3. Use `read_excel()` to read the c2015 dataset. Use function `class()` to check the type of data you just read in. You will notice that the data now is not just a data frame, it is also a tibble. A tibble is a generalization of a data frame

```
c2015 = read_excel("C:/Users/student/Documents/Senior Year/Fall Semester/R Analytics/Data Set/c2015.xlsx")
class(c2015)
```

```
## [1] "tbl_df"      "tbl"        "data.frame"
```

4. Use `dim` function to check the dimension of the data. Since this data is quite big, a common practice is to randomly subset the data to analyze

```
dim(c2015)
```

```
## [1] 80587    28
```

```
set.seed(2019)
```

```
sample2015 <- c2015[sample(nrow(c2015),1000),]
sample2015
```

```
## # A tibble: 1,000 x 28
```

```
##   STATE ST_CASE VEH_NO PER_NO COUNTY DAY MONTH HOUR MINUTE AGE SEX
##   <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <dbl> <dbl> <chr> <chr>
## 1 New ~   340336     1     1    27   19 Sept~     3    17 Unkn~ Unkn~
## 2 Ariz~    40327     1     1    13    7 May      22    15 47   Fema~
## 3 Tenn~   470789     1     1   163    2 Dece~     8    26 23   Male
## 4 Minn~   270119     2     4    59   16 May      21    59 15   Fema~
## 5 Miss~   290576     1     1   201    2 Octo~    15    38 55   Male
```

```
## 6 Cali~ 62865 1 1 19 6 June 15 20 56 Male
## 7 New ~ 330095 0 1 15 3 Dece~ 14 32 26 Male
## 8 Iowa 190173 0 1 127 30 Augu~ 20 20 63 Male
## 9 Cali~ 62263 2 4 13 17 Dece~ 7 41 6 Male
## 10 Alab~ 10286 5 1 115 30 May 14 36 32 Male
## # ... with 990 more rows, and 17 more variables: PER_TYP <chr>,
## # INJ_SEV <chr>, SEAT_POS <chr>, DRINKING <chr>, YEAR <dbl>,
## # MAN_COLL <chr>, OWNER <chr>, MOD_YEAR <chr>, TRAV_SP <chr>,
## # DEFORMED <chr>, DAY_WEEK <chr>, ROUTE <chr>, LATITUDE <dbl>,
## # LONGITUD <dbl>, HARM_EV <chr>, LGT_COND <chr>, WEATHER <chr>
```

5. Use summary function to have a quick look at the data. You will notice there is one variable is actually a constant. Remove that variable from the data.

```
summary(c2015)
```

```
##      STATE      ST_CASE      VEH_NO      PER_NO
## Length:80587      Min.   : 10001      Min.   : 0.000      Min.   : 1.000
## Class :character      1st Qu.:121977      1st Qu.: 1.000      1st Qu.: 1.000
## Mode  :character      Median :270282      Median : 1.000      Median : 1.000
##                               Mean  :275607      Mean  : 1.389      Mean  : 1.629
##                               3rd Qu.:420645      3rd Qu.: 2.000      3rd Qu.: 2.000
##                               Max.   :560130      Max.   :58.000      Max.   :51.000
##
##      COUNTY      DAY      MONTH      HOUR
## Min.   : 1.00      Min.   : 1.00      Length:80587      Min.   : 0.00
## 1st Qu.: 31.00      1st Qu.: 8.00      Class :character      1st Qu.: 8.00
## Median : 71.00      Median :15.00      Mode  :character      Median :15.00
## Mean   : 91.72      Mean   :15.52                               Mean  :14.01
## 3rd Qu.:115.00      3rd Qu.:23.00                               3rd Qu.:19.00
## Max.   :999.00      Max.   :31.00                               Max.   :99.00
##
##      MINUTE      AGE      SEX      PER_TYP
## Min.   : 0.00      Length:80587      Length:80587      Length:80587
## 1st Qu.:14.00      Class :character      Class :character      Class :character
## Median :29.00      Mode  :character      Mode  :character      Mode  :character
## Mean   :28.42
## 3rd Qu.:44.00
## Max.   :59.00
## NA's   :377
##      INJ_SEV      SEAT_POS      DRINKING      YEAR
## Length:80587      Length:80587      Length:80587      Min.   :2015
## Class :character      Class :character      Class :character      1st Qu.:2015
## Mode  :character      Mode  :character      Mode  :character      Median :2015
##                               Mean  :2015
##                               3rd Qu.:2015
##                               Max.   :2015
##
##      MAN_COLL      OWNER      MOD_YEAR
## Length:80587      Length:80587      Length:80587
## Class :character      Class :character      Class :character
## Mode  :character      Mode  :character      Mode  :character
##
```

```
##
##
##
##   TRAV_SP      DEFORMED      DAY_WEEK
## Length:80587   Length:80587   Length:80587
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
##
##
##
##   ROUTE      LATITUDE      LONGITUD      HARM_EV
## Length:80587 Min.   :19.27   Min.   :-166.55 Length:80587
## Class :character 1st Qu.:32.95   1st Qu.: -98.36 Class :character
## Mode  :character Median :36.21   Median : -87.90 Mode  :character
##                Mean  :36.52   Mean  : -92.43
##                3rd Qu.:40.56   3rd Qu.: -81.51
##                Max.   :65.18   Max.   : -67.60
##                NA's    :479     NA's    :479
##   LGT_COND      WEATHER
## Length:80587   Length:80587
## Class :character Class :character
## Mode  :character Mode  :character
##
##
##
##
```

6. Check the number of missing values (NA) in each column

```
null = is.na(c2015)

colSums(null)
```

```
##   STATE ST_CASE VEH_NO PER_NO COUNTY DAY MONTH HOUR
##     0      0      0      0      0      0      0      0
## MINUTE AGE     SEX PER_TYP INJ_SEV SEAT_POS DRINKING YEAR
##   377      0      0      0      0      0      0      0
## MAN_COLL OWNER MOD_YEAR TRAV_SP DEFORMED DAY_WEEK ROUTE LATITUDE
##   7197    7197    7197    7197    7197      0      0      479
## LONGITUD HARM_EV LGT_COND WEATHER
##     479      0      0      0
```

7 There are missing values in this data that are not NAs. Identify the form of these missing values. Check the number of these missing values in each column

```
sample2015=replace(sample2015,"unknow",NA)
sample2015=replace(sample2015,'Not Rep', NA)
sample2015
```

```
## # A tibble: 1,000 x 30
##   STATE ST_CASE VEH_NO PER_NO COUNTY DAY MONTH HOUR MINUTE AGE SEX
```

```
##      <chr>      <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <dbl> <dbl> <chr> <chr>
## 1 New ~ 340336      1      1      27      19 Sept~      3      17 Unkn~ Unkn~
## 2 Ariz~ 40327      1      1      13      7 May      22      15 47      Fema~
## 3 Tenn~ 470789     1      1     163      2 Dece~      8      26 23      Male
## 4 Minn~ 270119     2      4      59     16 May      21      59 15      Fema~
## 5 Miss~ 290576     1      1     201      2 Octo~     15      38 55      Male
## 6 Cali~ 62865      1      1      19      6 June      15      20 56      Male
## 7 New ~ 330095     0      1      15      3 Dece~     14      32 26      Male
## 8 Iowa 190173      0      1     127     30 Augu~     20      20 63      Male
## 9 Cali~ 62263      2      4      13     17 Dece~      7      41 6       Male
## 10 Alab~ 10286     5      1     115     30 May      14      36 32      Male
## # ... with 990 more rows, and 19 more variables: PER_TYP <chr>,
## #   INJ_SEV <chr>, SEAT_POS <chr>, DRINKING <chr>, YEAR <dbl>,
## #   MAN_COLL <chr>, OWNER <chr>, MOD_YEAR <chr>, TRAV_SP <chr>,
## #   DEFORMED <chr>, DAY_WEEK <chr>, ROUTE <chr>, LATITUDE <dbl>,
## #   LONGITUD <dbl>, HARM_EV <chr>, LGT_COND <chr>, WEATHER <chr>,
## #   unknow <lgl>, `Not Rep` <lgl>
```

8. Change the missing values in SEX variable to “Female”

```
sample2015$SEX = replace(sample2015$SEX, NA, "Female")
sample2015$SEX
```

```
##      [1] "Unknown" "Female"  "Male"    "Female"  "Male"    "Male"
##      [7] "Male"    "Male"    "Male"    "Male"    "Male"    "Male"
##     [13] "Female"  "Male"    "Female"  "Female"  "Male"    "Male"
##     [19] "Male"    "Female"  "Male"    "Male"    "Male"    "Male"
##     [25] "Male"    "Male"    "Male"    "Male"    "Male"    "Male"
##     [31] "Male"    "Male"    "Male"    "Male"    "Female"  "Male"
##     [37] "Female"  "Male"    "Male"    "Male"    "Female"  "Male"
##     [43] "Male"    "Male"    "Female"  "Male"    "Male"    "Male"
##     [49] "Male"    "Male"    "Male"    "Male"    "Male"    "Male"
##     [55] "Female"  "Female"  "Male"    "Male"    "Male"    "Male"
##     [61] "Male"    "Male"    "Male"    "Female"  "Female"  "Female"
##     [67] "Male"    "Male"    "Female"  "Female"  "Male"    "Male"
##     [73] "Female"  "Male"    "Male"    "Female"  "Male"    "Male"
##     [79] "Male"    "Female"  "Female"  "Male"    "Male"    "Female"
##     [85] "Male"    "Male"    "Male"    "Male"    "Female"  "Male"
##     [91] "Male"    "Male"    "Male"    "Male"    "Female"  "Female"
##     [97] "Male"    "Male"    "Male"    "Male"    "Female"  "Male"
##    [103] "Female"  "Male"    "Male"    "Male"    "Unknown" "Female"
##    [109] "Male"    "Female"  "Female"  "Male"    "Female"  "Male"
##    [115] "Male"    "Male"    "Male"    "Male"    "Male"    "Male"
##    [121] "Male"    "Male"    "Male"    "Female"  "Female"  "Female"
##    [127] "Female"  "Male"    "Male"    "Male"    "Male"    "Female"
##    [133] "Male"    "Male"    "Male"    "Male"    "Female"  "Female"
##    [139] "Male"    "Male"    "Female"  "Female"  "Female"  "Male"
##    [145] "Female"  "Female"  "Female"  "Male"    "Male"    "Male"
##    [151] "Female"  "Female"  "Male"    "Female"  "Male"    "Female"
##    [157] "Male"    "Female"  "Male"    "Female"  "Male"    "Not Rep"
##    [163] "Male"    "Female"  "Female"  "Female"  "Male"    "Male"
##    [169] "Male"    "Female"  "Male"    "Male"    "Male"    "Male"
##    [175] "Male"    "Female"  "Male"    "Female"  "Male"    "Female"
```

##	[181]	"Female"	"Male"	"Male"	"Female"	"Female"	"Female"
##	[187]	"Male"	"Male"	"Female"	"Male"	"Male"	"Female"
##	[193]	"Female"	"Male"	"Female"	"Male"	"Female"	"Male"
##	[199]	"Male"	"Female"	"Female"	"Male"	"Male"	"Male"
##	[205]	"Female"	"Male"	"Female"	"Female"	"Female"	"Male"
##	[211]	"Male"	"Female"	"Female"	"Female"	"Male"	"Female"
##	[217]	"Female"	"Male"	"Female"	"Female"	"Male"	"Female"
##	[223]	"Male"	"Female"	"Unknown"	"Male"	"Male"	"Female"
##	[229]	"Male"	"Male"	"Male"	"Male"	"Female"	"Male"
##	[235]	"Male"	"Female"	"Male"	"Male"	"Male"	"Female"
##	[241]	"Female"	"Male"	"Male"	"Male"	"Male"	"Male"
##	[247]	"Female"	"Male"	"Female"	"Male"	"Female"	"Male"
##	[253]	"Male"	"Male"	"Male"	"Female"	"Female"	"Female"
##	[259]	"Male"	"Male"	"Female"	"Male"	"Male"	"Female"
##	[265]	"Male"	"Male"	"Female"	"Male"	"Female"	"Male"
##	[271]	"Male"	"Female"	"Male"	"Male"	"Female"	"Female"
##	[277]	"Male"	"Female"	"Male"	"Male"	"Male"	"Male"
##	[283]	"Female"	"Female"	"Female"	"Female"	"Unknown"	"Female"
##	[289]	"Male"	"Male"	"Male"	"Male"	"Male"	"Female"
##	[295]	"Female"	"Male"	"Male"	"Female"	"Male"	"Female"
##	[301]	"Male"	"Male"	"Male"	"Male"	"Male"	"Male"
##	[307]	"Female"	"Male"	"Male"	"Male"	"Male"	"Male"
##	[313]	"Male"	"Male"	"Male"	"Male"	"Female"	"Male"
##	[319]	"Male"	"Male"	"Male"	"Female"	"Female"	"Male"
##	[325]	"Male"	"Male"	"Female"	"Female"	"Male"	"Male"
##	[331]	"Male"	"Male"	"Male"	"Male"	"Male"	"Male"
##	[337]	"Female"	"Female"	"Female"	"Male"	"Male"	"Male"
##	[343]	"Male"	"Male"	"Male"	"Male"	"Male"	"Male"
##	[349]	"Male"	"Male"	"Female"	"Male"	"Male"	"Male"
##	[355]	"Male"	"Male"	"Male"	"Male"	"Male"	"Female"
##	[361]	"Male"	"Male"	"Male"	"Male"	"Female"	"Female"
##	[367]	"Male"	"Male"	"Female"	"Male"	"Female"	"Female"
##	[373]	"Male"	"Male"	"Female"	"Female"	"Male"	"Male"
##	[379]	"Female"	"Male"	"Male"	"Female"	"Female"	"Male"
##	[385]	"Male"	"Male"	"Female"	"Male"	"Female"	"Unknown"
##	[391]	"Male"	"Male"	"Female"	"Female"	"Male"	"Female"
##	[397]	"Female"	"Female"	"Male"	"Male"	"Male"	"Male"
##	[403]	"Male"	"Female"	"Male"	"Male"	"Male"	"Male"
##	[409]	"Female"	"Female"	"Male"	"Male"	"Male"	"Male"
##	[415]	"Male"	"Male"	"Male"	"Female"	"Male"	"Male"
##	[421]	"Male"	"Female"	"Male"	"Male"	"Male"	"Female"
##	[427]	"Male"	"Male"	"Female"	"Male"	"Male"	"Female"
##	[433]	"Female"	"Male"	"Female"	"Male"	"Female"	"Male"
##	[439]	"Male"	"Male"	"Female"	"Female"	"Female"	"Male"
##	[445]	"Female"	"Male"	"Female"	"Male"	"Male"	"Female"
##	[451]	"Male"	"Male"	"Female"	"Female"	"Male"	"Male"
##	[457]	"Female"	"Male"	"Male"	"Female"	"Female"	"Male"
##	[463]	"Female"	"Male"	"Female"	"Male"	"Female"	"Female"
##	[469]	"Female"	"Male"	"Female"	"Male"	"Male"	"Male"
##	[475]	"Male"	"Female"	"Male"	"Male"	"Male"	"Male"
##	[481]	"Male"	"Male"	"Female"	"Male"	"Female"	"Male"
##	[487]	"Male"	"Female"	"Male"	"Female"	"Male"	"Male"
##	[493]	"Male"	"Female"	"Male"	"Female"	"Male"	"Female"
##	[499]	"Male"	"Female"	"Male"	"Female"	"Female"	"Female"

##	[505]	"Male"	"Male"	"Male"	"Male"	"Male"	"Female"
##	[511]	"Male"	"Male"	"Male"	"Male"	"Female"	"Female"
##	[517]	"Female"	"Male"	"Female"	"Female"	"Female"	"Male"
##	[523]	"Male"	"Male"	"Male"	"Female"	"Female"	"Male"
##	[529]	"Male"	"Male"	"Female"	"Female"	"Male"	"Male"
##	[535]	"Female"	"Male"	"Female"	"Male"	"Male"	"Male"
##	[541]	"Female"	"Male"	"Male"	"Male"	"Female"	"Male"
##	[547]	"Male"	"Male"	"Male"	"Male"	"Male"	"Female"
##	[553]	"Female"	"Male"	"Male"	"Male"	"Female"	"Female"
##	[559]	"Male"	"Male"	"Female"	"Female"	"Male"	"Female"
##	[565]	"Male"	"Female"	"Male"	"Male"	"Male"	"Male"
##	[571]	"Male"	"Male"	"Female"	"Male"	"Male"	"Male"
##	[577]	"Male"	"Female"	"Male"	"Female"	"Female"	"Male"
##	[583]	"Male"	"Female"	"Female"	"Male"	"Female"	"Male"
##	[589]	"Female"	"Male"	"Male"	"Male"	"Female"	"Female"
##	[595]	"Female"	"Male"	"Male"	"Male"	"Male"	"Female"
##	[601]	"Female"	"Female"	"Male"	"Male"	"Male"	"Male"
##	[607]	"Male"	"Male"	"Male"	"Female"	"Female"	"Male"
##	[613]	"Female"	"Male"	"Unknown"	"Male"	"Female"	"Male"
##	[619]	"Female"	"Male"	"Female"	"Female"	"Male"	"Female"
##	[625]	"Male"	"Male"	"Male"	"Male"	"Female"	"Female"
##	[631]	"Unknown"	"Male"	"Female"	"Male"	"Male"	"Female"
##	[637]	"Male"	"Female"	"Male"	"Male"	"Not Rep"	"Female"
##	[643]	"Male"	"Male"	"Male"	"Female"	"Female"	"Female"
##	[649]	"Male"	"Male"	"Male"	"Male"	"Male"	"Male"
##	[655]	"Female"	"Male"	"Female"	"Female"	"Male"	"Male"
##	[661]	"Female"	"Female"	"Male"	"Male"	"Male"	"Male"
##	[667]	"Female"	"Female"	"Male"	"Female"	"Male"	"Male"
##	[673]	"Male"	"Female"	"Male"	"Female"	"Male"	"Male"
##	[679]	"Male"	"Male"	"Male"	"Female"	"Male"	"Male"
##	[685]	"Female"	"Male"	"Male"	"Male"	"Female"	"Male"
##	[691]	"Male"	"Male"	"Male"	"Male"	"Male"	"Male"
##	[697]	"Male"	"Male"	"Male"	"Female"	"Female"	"Female"
##	[703]	"Male"	"Male"	"Male"	"Male"	"Male"	"Female"
##	[709]	"Male"	"Female"	"Male"	"Female"	"Male"	"Female"
##	[715]	"Male"	"Male"	"Male"	"Male"	"Male"	"Male"
##	[721]	"Female"	"Male"	"Male"	"Female"	"Male"	"Male"
##	[727]	"Female"	"Male"	"Male"	"Male"	"Male"	"Male"
##	[733]	"Male"	"Female"	"Male"	"Male"	"Female"	"Female"
##	[739]	"Male"	"Female"	"Female"	"Female"	"Male"	"Male"
##	[745]	"Male"	"Male"	"Male"	"Male"	"Female"	"Male"
##	[751]	"Male"	"Male"	"Male"	"Female"	"Male"	"Male"
##	[757]	"Male"	"Male"	"Male"	"Male"	"Male"	"Male"
##	[763]	"Male"	"Male"	"Male"	"Male"	"Male"	"Male"
##	[769]	"Female"	"Male"	"Male"	"Male"	"Male"	"Female"
##	[775]	"Male"	"Male"	"Male"	"Male"	"Female"	"Female"
##	[781]	"Male"	"Female"	"Male"	"Unknown"	"Female"	"Male"
##	[787]	"Male"	"Female"	"Male"	"Female"	"Male"	"Male"
##	[793]	"Male"	"Male"	"Female"	"Male"	"Female"	"Female"
##	[799]	"Female"	"Male"	"Female"	"Female"	"Female"	"Male"
##	[805]	"Female"	"Male"	"Male"	"Female"	"Female"	"Female"
##	[811]	"Female"	"Female"	"Female"	"Female"	"Female"	"Male"
##	[817]	"Female"	"Male"	"Male"	"Male"	"Male"	"Male"
##	[823]	"Male"	"Male"	"Female"	"Male"	"Male"	"Female"

```
## [829] "Male" "Male" "Female" "Female" "Male" "Male"
## [835] "Male" "Male" "Male" "Male" "Male" "Male"
## [841] "Male" "Female" "Male" "Female" "Male" "Female"
## [847] "Male" "Male" "Male" "Male" "Female" "Male"
## [853] "Male" "Male" "Male" "Male" "Female" "Male"
## [859] "Female" "Male" "Male" "Male" "Male" "Male"
## [865] "Male" "Male" "Male" "Male" "Male" "Male"
## [871] "Female" "Male" "Male" "Male" "Male" "Male"
## [877] "Female" "Male" "Male" "Female" "Male" "Male"
## [883] "Female" "Male" "Female" "Female" "Male" "Female"
## [889] "Male" "Male" "Male" "Male" "Female" "Male"
## [895] "Male" "Female" "Male" "Male" "Female" "Female"
## [901] "Female" "Male" "Male" "Male" "Male" "Male"
## [907] "Male" "Female" "Male" "Male" "Male" "Female"
## [913] "Male" "Male" "Male" "Male" "Female" "Female"
## [919] "Male" "Male" "Female" "Male" "Male" "Male"
## [925] "Male" "Male" "Male" "Male" "Male" "Male"
## [931] "Female" "Male" "Female" "Male" "Male" "Male"
## [937] "Male" "Male" "Female" "Male" "Male" "Male"
## [943] "Male" "Female" "Male" "Male" "Female" "Male"
## [949] "Male" "Male" "Male" "Female" "Female" "Male"
## [955] "Male" "Male" "Male" "Female" "Female" "Female"
## [961] "Male" "Male" "Male" "Unknown" "Male" "Male"
## [967] "Male" "Male" "Female" "Male" "Male" "Female"
## [973] "Male" "Female" "Female" "Female" "Male" "Male"
## [979] "Male" "Male" "Male" "Male" "Male" "Male"
## [985] "Male" "Male" "Male" "Male" "Female" "Male"
## [991] "Male" "Female" "Female" "Male" "Female" "Male"
## [997] "Female" "Male" "Female" "Female"
```

9. Fix the AGE variable so that it is in the right form and has no missing values.

```
sample2015['AGE'] [sample2015['AGE']== "Less than 1"]<- "0"
sample2015$AGE=as.numeric(sample2015$AGE)
```

```
## Warning: NAs introduced by coercion
```

```
sample2015$AGE[is.na(sample2015$AGE)]<- mean (sample2015$AGE,na.rm = TRUE)
sample2015$AGE
```

```
## [1] 39.28252 47.00000 23.00000 15.00000 55.00000 56.00000 26.00000
## [8] 63.00000 6.00000 32.00000 54.00000 44.00000 10.00000 79.00000
## [15] 39.00000 15.00000 44.00000 63.00000 56.00000 50.00000 43.00000
## [22] 79.00000 53.00000 71.00000 42.00000 35.00000 22.00000 59.00000
## [29] 28.00000 61.00000 78.00000 54.00000 84.00000 39.28252 31.00000
## [36] 29.00000 12.00000 47.00000 21.00000 58.00000 51.00000 34.00000
## [43] 65.00000 24.00000 64.00000 20.00000 31.00000 29.00000 34.00000
## [50] 31.00000 38.00000 25.00000 20.00000 60.00000 39.28252 19.00000
## [57] 31.00000 24.00000 16.00000 43.00000 43.00000 59.00000 29.00000
## [64] 50.00000 62.00000 38.00000 78.00000 24.00000 36.00000 65.00000
## [71] 58.00000 18.00000 39.28252 8.00000 31.00000 21.00000 45.00000
## [78] 21.00000 76.00000 77.00000 56.00000 25.00000 73.00000 35.00000
```

```

## [85] 51.00000 33.00000 55.00000 19.00000 5.00000 37.00000 20.00000
## [92] 17.00000 32.00000 30.00000 22.00000 46.00000 36.00000 30.00000
## [99] 89.00000 56.00000 68.00000 54.00000 18.00000 62.00000 24.00000
## [106] 83.00000 57.00000 24.00000 57.00000 12.00000 44.00000 18.00000
## [113] 22.00000 39.00000 23.00000 39.00000 35.00000 19.00000 33.00000
## [120] 9.00000 31.00000 20.00000 30.00000 89.00000 83.00000 3.00000
## [127] 19.00000 1.00000 61.00000 31.00000 49.00000 82.00000 39.00000
## [134] 51.00000 66.00000 76.00000 52.00000 20.00000 72.00000 29.00000
## [141] 35.00000 27.00000 79.00000 52.00000 46.00000 57.00000 17.00000
## [148] 58.00000 57.00000 15.00000 61.00000 41.00000 31.00000 56.00000
## [155] 38.00000 44.00000 57.00000 59.00000 45.00000 41.00000 32.00000
## [162] 3.00000 26.00000 24.00000 42.00000 20.00000 53.00000 23.00000
## [169] 39.00000 26.00000 32.00000 44.00000 11.00000 37.00000 26.00000
## [176] 17.00000 43.00000 27.00000 32.00000 90.00000 17.00000 28.00000
## [183] 35.00000 8.00000 48.00000 13.00000 35.00000 27.00000 51.00000
## [190] 91.00000 29.00000 65.00000 26.00000 16.00000 37.00000 75.00000
## [197] 26.00000 22.00000 56.00000 32.00000 45.00000 48.00000 58.00000
## [204] 21.00000 49.00000 27.00000 10.00000 17.00000 15.00000 47.00000
## [211] 21.00000 1.00000 24.00000 27.00000 10.00000 22.00000 52.00000
## [218] 56.00000 55.00000 53.00000 2.00000 59.00000 27.00000 46.00000
## [225] 39.28252 41.00000 62.00000 3.00000 27.00000 22.00000 40.00000
## [232] 64.00000 36.00000 38.00000 16.00000 58.00000 89.00000 64.00000
## [239] 18.00000 20.00000 37.00000 74.00000 50.00000 25.00000 42.00000
## [246] 62.00000 19.00000 61.00000 67.00000 24.00000 48.00000 50.00000
## [253] 47.00000 33.00000 75.00000 67.00000 95.00000 39.28252 19.00000
## [260] 36.00000 38.00000 84.00000 64.00000 58.00000 21.00000 21.00000
## [267] 54.00000 67.00000 8.00000 22.00000 63.00000 26.00000 67.00000
## [274] 53.00000 12.00000 41.00000 80.00000 19.00000 26.00000 26.00000
## [281] 36.00000 87.00000 21.00000 46.00000 25.00000 43.00000 39.28252
## [288] 24.00000 47.00000 42.00000 35.00000 53.00000 50.00000 34.00000
## [295] 5.00000 45.00000 28.00000 29.00000 38.00000 60.00000 10.00000
## [302] 79.00000 34.00000 28.00000 19.00000 35.00000 81.00000 54.00000
## [309] 40.00000 47.00000 53.00000 61.00000 19.00000 67.00000 8.00000
## [316] 24.00000 24.00000 0.00000 43.00000 16.00000 81.00000 16.00000
## [323] 44.00000 18.00000 27.00000 41.00000 69.00000 23.00000 56.00000
## [330] 72.00000 24.00000 17.00000 54.00000 60.00000 39.00000 25.00000
## [337] 26.00000 49.00000 7.00000 40.00000 39.28252 19.00000 25.00000
## [344] 53.00000 49.00000 51.00000 44.00000 51.00000 62.00000 31.00000
## [351] 15.00000 43.00000 68.00000 69.00000 34.00000 26.00000 40.00000
## [358] 15.00000 27.00000 11.00000 52.00000 22.00000 80.00000 58.00000
## [365] 46.00000 25.00000 35.00000 64.00000 10.00000 31.00000 7.00000
## [372] 16.00000 15.00000 31.00000 33.00000 20.00000 21.00000 69.00000
## [379] 25.00000 31.00000 34.00000 83.00000 27.00000 22.00000 33.00000
## [386] 53.00000 1.00000 24.00000 37.00000 39.28252 27.00000 19.00000
## [393] 39.28252 13.00000 21.00000 24.00000 54.00000 3.00000 47.00000
## [400] 35.00000 30.00000 21.00000 55.00000 30.00000 39.00000 38.00000
## [407] 35.00000 35.00000 74.00000 20.00000 48.00000 33.00000 43.00000
## [414] 22.00000 53.00000 49.00000 85.00000 17.00000 33.00000 72.00000
## [421] 33.00000 56.00000 69.00000 29.00000 29.00000 75.00000 40.00000
## [428] 36.00000 55.00000 37.00000 60.00000 59.00000 47.00000 51.00000
## [435] 18.00000 18.00000 61.00000 31.00000 50.00000 29.00000 25.00000
## [442] 38.00000 13.00000 22.00000 17.00000 25.00000 39.00000 39.00000
## [449] 57.00000 16.00000 89.00000 24.00000 34.00000 48.00000 32.00000
## [456] 49.00000 36.00000 4.00000 9.00000 23.00000 57.00000 62.00000

```



```

## [463] 57.00000 1.00000 21.00000 62.00000 77.00000 51.00000 24.00000
## [470] 59.00000 10.00000 52.00000 37.00000 55.00000 39.28252 69.00000
## [477] 27.00000 25.00000 60.00000 29.00000 15.00000 27.00000 14.00000
## [484] 25.00000 53.00000 23.00000 26.00000 57.00000 55.00000 18.00000
## [491] 80.00000 82.00000 21.00000 14.00000 63.00000 28.00000 21.00000
## [498] 23.00000 66.00000 60.00000 44.00000 51.00000 20.00000 61.00000
## [505] 16.00000 35.00000 18.00000 41.00000 83.00000 74.00000 27.00000
## [512] 25.00000 31.00000 71.00000 56.00000 80.00000 28.00000 23.00000
## [519] 20.00000 32.00000 12.00000 27.00000 62.00000 29.00000 59.00000
## [526] 21.00000 31.00000 23.00000 25.00000 45.00000 54.00000 20.00000
## [533] 6.00000 33.00000 89.00000 36.00000 54.00000 51.00000 71.00000
## [540] 43.00000 36.00000 47.00000 81.00000 93.00000 24.00000 26.00000
## [547] 9.00000 36.00000 58.00000 56.00000 35.00000 46.00000 68.00000
## [554] 31.00000 37.00000 51.00000 20.00000 60.00000 85.00000 19.00000
## [561] 5.00000 57.00000 48.00000 78.00000 22.00000 67.00000 63.00000
## [568] 34.00000 47.00000 20.00000 6.00000 26.00000 0.00000 37.00000
## [575] 37.00000 27.00000 60.00000 61.00000 17.00000 16.00000 51.00000
## [582] 15.00000 48.00000 0.00000 78.00000 33.00000 25.00000 28.00000
## [589] 35.00000 26.00000 42.00000 20.00000 23.00000 23.00000 2.00000
## [596] 21.00000 54.00000 20.00000 10.00000 48.00000 42.00000 25.00000
## [603] 50.00000 21.00000 40.00000 44.00000 2.00000 5.00000 54.00000
## [610] 14.00000 27.00000 46.00000 62.00000 63.00000 39.28252 53.00000
## [617] 39.00000 61.00000 53.00000 22.00000 26.00000 59.00000 22.00000
## [624] 22.00000 86.00000 2.00000 35.00000 64.00000 15.00000 43.00000
## [631] 39.28252 35.00000 61.00000 53.00000 50.00000 34.00000 24.00000
## [638] 76.00000 15.00000 62.00000 39.28252 49.00000 17.00000 48.00000
## [645] 29.00000 93.00000 20.00000 65.00000 49.00000 34.00000 57.00000
## [652] 18.00000 26.00000 38.00000 40.00000 90.00000 51.00000 57.00000
## [659] 27.00000 44.00000 22.00000 64.00000 42.00000 51.00000 60.00000
## [666] 9.00000 27.00000 44.00000 60.00000 38.00000 22.00000 61.00000
## [673] 10.00000 18.00000 54.00000 82.00000 35.00000 37.00000 46.00000
## [680] 49.00000 15.00000 21.00000 24.00000 20.00000 25.00000 78.00000
## [687] 29.00000 21.00000 21.00000 63.00000 33.00000 22.00000 63.00000
## [694] 2.00000 40.00000 52.00000 41.00000 3.00000 28.00000 89.00000
## [701] 27.00000 30.00000 29.00000 18.00000 49.00000 2.00000 22.00000
## [708] 43.00000 65.00000 78.00000 23.00000 43.00000 40.00000 23.00000
## [715] 34.00000 83.00000 42.00000 27.00000 38.00000 51.00000 3.00000
## [722] 76.00000 59.00000 11.00000 55.00000 22.00000 37.00000 52.00000
## [729] 65.00000 76.00000 35.00000 49.00000 58.00000 65.00000 18.00000
## [736] 56.00000 67.00000 2.00000 60.00000 23.00000 82.00000 37.00000
## [743] 18.00000 23.00000 23.00000 18.00000 20.00000 23.00000 67.00000
## [750] 61.00000 35.00000 23.00000 39.00000 17.00000 48.00000 6.00000
## [757] 13.00000 63.00000 31.00000 29.00000 70.00000 39.00000 30.00000
## [764] 73.00000 19.00000 46.00000 71.00000 2.00000 50.00000 77.00000
## [771] 57.00000 45.00000 51.00000 45.00000 71.00000 37.00000 22.00000
## [778] 20.00000 50.00000 35.00000 33.00000 80.00000 59.00000 39.28252
## [785] 17.00000 25.00000 21.00000 89.00000 59.00000 39.00000 56.00000
## [792] 44.00000 21.00000 21.00000 27.00000 27.00000 21.00000 20.00000
## [799] 28.00000 67.00000 41.00000 5.00000 64.00000 49.00000 58.00000
## [806] 32.00000 55.00000 29.00000 61.00000 44.00000 30.00000 31.00000
## [813] 27.00000 19.00000 57.00000 33.00000 19.00000 8.00000 24.00000
## [820] 27.00000 35.00000 21.00000 36.00000 23.00000 27.00000 63.00000
## [827] 26.00000 49.00000 50.00000 56.00000 33.00000 47.00000 31.00000
## [834] 60.00000 44.00000 67.00000 38.00000 29.00000 48.00000 54.00000

```

```
## [841] 13.00000 45.00000 58.00000 21.00000 36.00000 14.00000 66.00000
## [848] 19.00000 20.00000 62.00000 37.00000 66.00000 33.00000 21.00000
## [855] 30.00000 74.00000 22.00000 40.00000 25.00000 56.00000 19.00000
## [862] 14.00000 35.00000 23.00000 22.00000 53.00000 1.00000 2.00000
## [869] 22.00000 1.00000 24.00000 44.00000 30.00000 36.00000 29.00000
## [876] 57.00000 16.00000 65.00000 48.00000 49.00000 41.00000 58.00000
## [883] 6.00000 62.00000 15.00000 24.00000 60.00000 64.00000 19.00000
## [890] 19.00000 50.00000 40.00000 29.00000 68.00000 74.00000 23.00000
## [897] 53.00000 19.00000 23.00000 5.00000 29.00000 57.00000 22.00000
## [904] 21.00000 44.00000 18.00000 45.00000 16.00000 27.00000 46.00000
## [911] 44.00000 56.00000 25.00000 68.00000 21.00000 0.00000 63.00000
## [918] 57.00000 27.00000 25.00000 34.00000 18.00000 39.00000 62.00000
## [925] 34.00000 67.00000 59.00000 25.00000 40.00000 71.00000 7.00000
## [932] 14.00000 25.00000 43.00000 44.00000 19.00000 31.00000 30.00000
## [939] 85.00000 28.00000 29.00000 70.00000 13.00000 42.00000 51.00000
## [946] 24.00000 42.00000 28.00000 18.00000 25.00000 94.00000 1.00000
## [953] 22.00000 41.00000 26.00000 61.00000 17.00000 40.00000 28.00000
## [960] 65.00000 46.00000 41.00000 20.00000 39.28252 61.00000 40.00000
## [967] 27.00000 74.00000 52.00000 32.00000 18.00000 19.00000 37.00000
## [974] 66.00000 65.00000 71.00000 65.00000 54.00000 26.00000 60.00000
## [981] 55.00000 26.00000 37.00000 46.00000 43.00000 84.00000 31.00000
## [988] 47.00000 49.00000 35.00000 51.00000 22.00000 70.00000 48.00000
## [995] 73.00000 9.00000 76.00000 42.00000 29.00000 0.00000
```

10. Put the TRAV_SP(Travel Speed) variable in the right form (type) and remove all missing values. Calculate the average speed. You can use a non-base R function for this question. Hint: check out the function str_replace

```
sample2015$TRAV_SP <- substr(sample2015$TRAV_SP, 1, nchar(sample2015$TRAV_SP)-4)
sample2015$TRAV_SP <- as.numeric(as.character(sample2015$TRAV_SP))
```

```
## Warning: NAs introduced by coercion
```

```
here <- sample2015[!is.na(sample2015$TRAV_SP),]
mean(here$TRAV_SP)
```

```
## [1] 50.77188
```

11 Compare the average speed of those who had “No Apparent Injury” and the rest. What do you observe?

```
notinjured <- here[here$INJ_SEV == 'No Apparent Injury (0)',]
injurey <- here[here$INJ_SEV != 'No Apparent Injury (0)',]
mean(notinjured$TRAV_SP, na.rm = TRUE)
```

```
## [1] 44.63636
```

```
mean(injurey$TRAV_SP, na.rm=TRUE)
```

```
## [1] 53.09914
```

```
# Travel speed with injury is higher
```

12. Use the SEAT_POS variable to filter the data so that there is only drivers in the dataset. Compare the average speed of man drivers and woman drivers. Comment on the results

```
driver <- here[here$SEAT_POS == "Front Seat, Left Side",]  
maledriver <- driver[driver$SEX == 'Male',]  
femaledriver <- driver[driver$SEX == 'Female',]  
mean(maledriver$TRAV_SP, na.rm = TRUE)
```

```
## [1] 51.65333
```

```
mean(femaledriver$TRAV_SP, na.rm=TRUE)
```

```
## [1] 46.07407
```

```
#males in accident driver faster than females
```

Compare the average speed of drivers who drink and those who do not. Comment on the results. Hint: This calculation can be done manually or by using the aggregate function or by function in base R. For example:

```
drink <- driver[driver$DRINKING == 'Yes (Alcohol Involved)',]  
notdrunk <- driver[driver$DRINKING == 'No (Alcohol Not Involved)',]  
mean(drink$TRAV_SP, na.rm = TRUE)
```

```
## [1] 68.25
```

```
mean(notdrunk$TRAV_SP, na.rm = TRUE)
```

```
## [1] 44.94074
```

```
#people who drink drive faster than people who dont
```

14 Hypothesize about the age range of drivers who may drive more aggressively. Test your hypothesis by comparing the average speed of those in this age range and the rest. Comment on the results.

```
# I hypothesis that people under 25 drive faster than over 25
```

```
under <- driver[driver$AGE < 25,]  
over <- driver[driver$AGE >= 25,]  
mean(under$TRAV_SP, na.rm = TRUE)
```

```
## [1] 56.25641
```

```
mean (over$TRAV_SP, na.rm = TRUE)
```

```
## [1] 48.52381
```

```
#The results show that poeple under 25 tend to drive faster
```

#####15 If the data did not confirm your hypothesis in 14. Could you identify an age group of drivers who may drive more aggressively?

```
# problem 14 comfirmed my hypothesis so i did not do this step
```