CMTH642 Capstone

Geoffrey Clark

May 28, 2018

Fire Incident Data Overview

```
## Incidents
# Load & Combine Dataset from Yearwise .csv


I <- read.csv("../csv/2011_i.csv", header=T, stringsAsFactors = F, na.strings=c("","NA", " "))
# importSchema = c("character", "factor", "POSIXct", "POSIXct", "factor", "factor", "factor", "factor",


# metadata <- data.frame('year'=c(2011), 'nrow'=nrow(I))


for(yr in (2012:2016)){
  i_file <- paste("../csv/",yr, "_i.csv", sep="")

  # I <- rbind(I, read.csv(i_file, header=T))
  i_df <- read.csv(i_file, header=T, stringsAsFactors = F)
  # metadata <- rbind(metadata, c(yr, nrow(i_df)))
  I <- rbind(I, i_df)
  i_df <- NULL # keeps the environment tidy
}


# I$ARRIVE_DATE <- as.POSIXct(I$ARRIVE_DATE)
# str(I)

# sapply(I, summary)
```
```
## Responding Units
# Load & Combine Dataset from Yearwise .csv

RU <- read.csv("../csv/2011_ru.csv", header=T, stringsAsFactors = F, na.strings=c("","NA"))

ru_metadata <- data.frame('year'=c(2011), 'nrow'=nrow(RU))

for(yr in (2012:2016)){
  r_file <- paste("../csv/",yr, "_ru.csv", sep="")

  r_df <- read.csv(r_file, header=T, stringsAsFactors = F)
  ru_metadata <- rbind(ru_metadata, c(yr, nrow(r_df)))
  RU <- rbind(RU, r_df)
  r_df <- NULL # keeps the environment tidy
}
```
```
## Dates
# In this section I work with the three date features: DISPATCH_DATE, ARRIVE_DATE & INCIDENT_DATE
# In fact, DISPATCH_DATE & ARRIVE_DATE are identical so I drop one (arbitrary choice: ARRIVE_DATE)
```

```r
# This is some of the initial exploring of the dates I did before noticing that DISPATCH & ARRIVE dates
# I've left the code here for observation. It's commented out to save computation time & resources

# nrow(I[!grepl("\\d{4}-\\d{2}-\\d{2} \\d{2}:\\d{2}:\\d{2}", I$DISPATCH_DATE),]) #16802
# nrow(I[!grepl("\\d{4}-\\d{2}-\\d{2} \\d{2}:\\d{2}:\\d{2}", I$ARRIVE_DATE),]) #16802
# nrow(I[!grepl("\\d{4}-\\d{2}-\\d{2} \\d{2}:\\d{2}:\\d{2}", I$ARRIVE_DATE) & !grepl("\\d{4}-\\d{2}-\\d

# Originally I was reading characters in from the .csv as factors. I changed this to reduce overhead.
# I$DISPATCH_DATE <- as.character(I$DISPATCH_DATE)
# I$ARRIVE_DATE <- as.character(I$ARRIVE_DATE)
# I$INCINCIDENT_DATE <- as.character(I$INCIDENT_DATE)

# I[,c('DISPATCH_DATE','ARRIVE_DATE')] <- as.character(I[,c('DISPATCH_DATE','ARRIVE_DATE')]) # Slow

# I[!grepl("\\d{4}-\\d{2}-\\d{2} \\d{2}:\\d{2}:\\d{2}", I$ARRIVE_DATE) & !grepl("\\d{4}-\\d{2}-\\d{2} \
# I[!grepl("\\d{4}-\\d{2}-\\d{2} \\d{2}:\\d{2}:\\d{2}", I$ARRIVE_DATE) & !grepl("\\d{4}-\\d{2}-\\d{2} \
# I[!grepl("\\d{2}/\\d{2}/\\d{4}", I$INCIDENT_DATE),'INCIDENT_DATE'] <- NA # Everyone obs. has a value

# nrow(I[!grepl("\\d{2}/\\d{2}/\\d{4}", I$INCIDENT_DATE),]) # 0


if(identical(I$DISPATCH_DATE, I$ARRIVE_DATE)){ I$ARRIVE_DATE <- NULL }

I[!grepl("\\d{4}-\\d{2}-\\d{2} \\d{2}:\\d{2}:\\d{2}", I$DISPATCH_DATE),'DISPATCH_DATE'] <- NA
# Cannot set to date type because of 16,802 Missing Values
I$DISPATCH_DATE <- as.POSIXct(I$DISPATCH_DATE)
# I$ARRIVE_DATE <- as.POSIXct(I$ARRIVE_DATE) # I set I$ARRIVE_DATE to null, above
I$INCIDENT_DATE <- as.POSIXct(I$INCIDENT_DATE, format="%d/%m/%Y")
# td <- as.Date(I$INCIDENT_DATE, format="%d/%m/%Y")

sapply(I[,c('DISPATCH_DATE','INCIDENT_DATE')], function(x) sum(is.na(x)))
```

```
## DISPATCH_DATE INCIDENT_DATE
##         16802             0
```

```r
# sapply(I, function(x) sum(is.na(x))/nrow(I))
```

Factors

```r
# This was a tricky part of the data prep to handle. This is a relatively sparse dataset (see below for
# but also has a lot of factors. I decided to keep the NAs as a level in the factors to illustrate rest
# in the analysis introduced by such sparse data.
factorSchema <- c("EVENT_TYPE","EVENT_TYPE_CD","MAIN_STREET","CROSS_STREET","FSA","ALARM_TO_FD","RESPONS

for(ft in factorSchema){
  # cat(ft, "\n")
  I[,ft] <- addNA(I[,ft]) # creates factors with NA level
}

# This sapply returns column NAs as percentage of total
# but doesn't format the output as nicely as the below function which I kept.
#
# sapply(I, function(x){
```

```
#   per <- sum(is.na(x))/nrow(I)
#   if(per * 100 > 0.001) return(per*100)
#   return(0)
# })


# outputs number of NAs per column as a percent
# format & round because big floats being returned & hard to read.
sapply(I, function(x){

  p <- sum(is.na(x))/nrow(I)
  format(round(p*100, 2), nsmall=2)
})
```

```
##                          INCIDENT_NUMBER
##                                   "0.00"
##                               EVENT_TYPE
##                                   "0.00"
##                            DISPATCH_DATE
##                                   "2.33"
##                            EVENT_TYPE_CD
##                                   "0.00"
##                              MAIN_STREET
##                                   "0.00"
##                             CROSS_STREET
##                                   "0.00"
##                                      FSA
##                                   "0.00"
##                          RESPONDING_UNITS
##                                   "0.00"
##                               FD_STATION
##                                   "0.00"
##              OFM_INVESTIGATIONS_CONTACTED
##                                   "0.00"
##                   AID_TO_FROM_OTHER_DEPTS
##                                   "0.00"
##                            INCIDENT_DATE
##                                   "0.00"
##                        INITIAL_CALL_HOUR
##                                   "0.00"
##                         INITIAL_CALL_MIN
##                                   "0.00"
##                         INITIAL_CALL_SEC
##                                   "0.00"
##                            DISPATCH_HOUR
##                                   "0.02"
##                             DISPATCH_MIN
##                                   "0.02"
##                             DISPATCH_SEC
##                                   "0.02"
##                             ONSCENE_HOUR
##                                   "2.13"
##                              ONSCENE_MIN
##                                   "2.13"
```

```
##                         ONSCENE_SEC
##                             "2.13"
##              INITIAL_UNIT_PERSONNEL
##                             "0.00"
##                 TOTAL_NUM_PERSONNEL
##                             "0.00"
##                             EST_KM
##                             "0.00"
##                        ALARM_TO_FD
##                             "0.00"
##                      RESPONSE_TYPE
##                             "0.00"
##                            RESCUES
##                             "0.00"
##                        FF_INJURIES
##                             "0.00"
##                      FF_FATALITIES
##                             "0.00"
##                     AGENT_APP_HOUR
##                            "98.29"
##                      AGENT_APP_MIN
##                            "98.29"
##                      AGENT_APP_SEC
##                            "98.29"
##                       CONTROL_DATE
##                            "19.89"
##                       CONTROL_HOUR
##                            "98.26"
##                        CONTROL_MIN
##                            "98.26"
##                        CONTROL_SEC
##                            "98.26"
##                   STATUS_ON_ARRIVAL
##                             "0.00"
##                              WATER
##                             "0.00"
##                       FIRE_CONTROL
##                             "0.00"
##                           PROPERTY
##                             "0.00"
##                     AREA_OF_ORIGIN
##                             "0.00"
##                    IGNITION_SOURCE
##                             "0.00"
##            FUEL_OF_IGNITION_SOURCE
##                             "0.00"
##      OBJECT_OR_MATERIAL_FIRST_IGNITED
##                             "0.00"
##                     POSSIBLE_CAUSE
##                             "0.00"
##                        VEH_PURPOSE
##                             "0.00"
##                           VEH_FUEL
##                             "0.00"
```

```
##                            EST_LOSS
##                             "0.00"
##                  INSURANCE_ESTIMATE
##                             "0.00"
##                   EST_VALUE_AT_RISK
##                             "0.00"
##                CIVILIAN_FIRE_INJURY
##                             "0.00"
##              CIVILIAN_FIRE_FATALITY
##                             "0.00"
##                    RESCUED_CHILDREN
##                             "0.00"
##                      RESCUED_ADULTS
##                             "0.00"
##                     RESCUED_SENIORS
##                             "0.00"
##                 PHYSICAL_CONDITION_1
##                             "0.00"
##                 PHYSICAL_CONDITION_2
##                             "0.00"
##                 PHYSICAL_CONDITION_3
##                             "0.00"
##                     CIV_FIRE_CONTROL
##                             "0.00"
##                       CIV_EVACUATION
##                             "0.00"
##              CIV_EVACUATION_REASON_1
##                             "0.00"
##              CIV_EVACUATION_REASON_2
##                             "0.00"
##                                 OPP
##                             "0.00"
##                                 MOE
##                             "0.00"
##                                TSSA
##                             "0.00"
##                                 ESA
##                             "0.00"
##                                 MOL
##                             "0.00"
##                                 EMS
##                             "0.00"
##                             CANUTEC
##                             "0.00"
##                                 GAS
##                             "0.00"
##                               HYDRO
##                             "0.00"
##            MUNICIPAL_BUILDING_OFFICE
##                             "0.00"
##             MUNICIPAL_HEALTH_OFFICE
##                             "0.00"
##                    MUNICIPAL_POLICE
##                             "0.00"
```

5

```
##                                                   OTHER
##                                                  "0.00"
##                                       INITIAL_DETECTION
##                                                  "0.00"
##                                              EXTENT_FIRE
##                                                  "0.00"
##                                             EXTENT_SMOKE
##                                                  "0.00"
##                                EST_NUM_PERSONS_DISPLACED
##                                                  "0.00"
##                                 POSSIBLE_BUSINESS_IMPACT
##                                                  "0.00"
##                                                  COMPLEX
##                                                 "98.78"
##                                               OCC_STATUS
##                                                  "0.00"
##                                                 OCC_TYPE
##                                                  "0.00"
##                                               BLD_STATUS
##                                                  "0.00"
##                                               BLD_HEIGHT
##                                                  "0.00"
##                                          LEVEL_OF_ORIGIN
##                                                  "0.00"
##                                          AGE_OF_STRUCTURE
##                                                  "0.00"
## SMOKE_ALARM_PRESENCE_AND_OPERATION_MAIN_FLOOR
##                                                  "0.00"
##                            SMOKE_ALARM_FAILURE_TO_OPERATE
##                                                  "0.00"
##                                         SMOKE_ALARM_TYPE
##                                                  "0.00"
##                            SMOKE_ALARM_OTHER_FLOOR_PRESENCE
##                                                  "0.00"
##                                SMOKE_ALARM_ON_ALL_FLOORS
##                                                  "0.00"
##                                SMOKE_ALARM_IMPACT_ON_EVAC
##                                                  "0.00"
##                            SMOKE_ALARM_IMPACT_ON_NUM_EVAC
##                                                  "0.00"
##                                FIRE_ALARM_SYSTEM_PRESENCE
##                                                  "0.00"
##                               FIRE_ALARM_SYSTEM_OPERATION
##                                                  "0.00"
##                                  FIRE_ALARM_SYSTEM_IMPACT
##                                                  "0.00"
##                                SPRINKLER_SYSTEM_PRESENCE
##                                                  "0.00"
##                               SPRINKLER_SYSTEM_ACTIVATION
##                                                  "0.00"
# sapply(I, function(x) sum(is.na(x)))
# apply(I, 2, function(x) sum(is.na(x))/nrow(I))
# apply(I, 1, function(x) sum(is.na(x))/ncol(I))
```

```
I$ROW_NAS <- apply(I, 1, function(x) sum(is.na(x)))

summary(I$ROW_NAS)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   12.00   58.00   58.00   58.22   58.00   70.00
# https://www.stat.berkeley.edu/~s133/dates.html
# Some cool temporal breakdowns
table(format(I$INCIDENT_DATE, '%Y'))

##
##   2011   2012   2013   2014   2015   2016
## 145365 120545 109576 111794 115664 117426
table(format(I$INCIDENT_DATE, '%A'))

##
##     Friday     Monday   Saturday     Sunday   Thursday   Tuesday Wednesday
##     105371     102893     104100     101854     102353     102102    101697
table(format(I$INCIDENT_DATE, '%B'))

##
##      April     August   December   February    January       July       June
##      57519      60069      64076      55778      63457      64209      60475
##      March        May   November    October  September
##      58605      63155      55662      59013      58352
table(format(I$DISPATCH_DATE, '%H'))

##
##    00    01    02    03    04    05    06    07    08    09    10    11
## 23611 20556 18501 16314 13644 13335 16246 20560 27819 32951 35511 36931
##    12    13    14    15    16    17    18    19    20    21    22    23
## 37595 37986 37090 36732 37350 38186 39041 37203 35225 33691 30588 26902
table(I$PROPERTY)

##
##            101    102    103    104    111    112    113    114    115
##   1375    877    534    251     55    254    115    850    200     81
##    121    122    123    124    125    126    127    128    131    132
##    105     81   1728   1472    561    162    306     80    373   5801
##    133    134    135    136    137    138    141    142    143    144
##    955   3504    384    275   3377   1383    153   1395    794   9504
##    145    151    152    153    154    155    156    157    158    161
##    114   6074   2934   3539    277    427    516    532    374    460
##    162    163    171    172    173    174    175    199    201    202
##    293    242     20    280    326     68     37    825    796     24
##    203    204    205    206    211    212    216    217    218    219
##     59    999     35    812    203   7672   5398     85    478    117
##    222    223    229    231    232    233    234    299    301    302
##    101  23119   1943   3030    925   1749   5528    463 122844  26743
##    303    311    321    322    323    331    332    333    334    335
##  21600   2691   7420   7509 204952   8944    346    308   1059    629
##    336    341    342    343    344    345    355    356    361    363
```

```
##   293    30     4     8    19     2  4867    82   798    81
##   364   365   366   367   368   369   399   401   402   403
##    15   208    52  1998     4    50   214  1706   156   603
##   404   405   406   407   408   409   410   411   412   413
##   293 13091   192  4689   136    12    54   718   146    45
##   414   415   416   417   418   498   499   501   502   503
##    53   159    33    14    23   860  3932  4049  2840   685
##   504   505   506   507   510   511   512   521   522   523
##   645   362    40   134  5218  1591    21   636    35   303
##   524   525   526   527   528   529   530   531   532   533
##    17   154   241    34    16   120   923    17    45    29
##   534   535   536   537   538   539   541   543   599   601
##    29   150     6    38    16   126     2   403  1485   431
##   602   603   604   605   606   607   609   611   612   613
##   647   743    85    63     1   130   106    21   176     2
##   614   615   616   617   618   619   620   621   622   623
##    97    30    25    65    16   216    24    18   171   175
##   624   625   626   627   629   631   632   633   634   635
##    54   288    85    25   191   251    96   397   117    17
##   636   637   638   639   640   641   642   643   644   645
##     2    18   210   227    43    84     8    10    76    27
##   646   649   651   652   653   654   655   659   661   662
##    29    74   260    41   343   130    81   331   125     7
##   663   664   665   669   671   672   673   674   675   676
##     1    45    16    27    54   192    23     8     9    17
##   677   678   679   681   682   683   684   685   686   687
##     5    38   443     8    41    15    12    76     4    14
##   689   690   691   699   701   702   703   704   705   706
##    14     5     6    23    74    40    36     6    12     3
##   707   708   709   710   711   712   714   715   716   717
##    11     7    76     5     1     1     3     1    56     2
##   719   721   722   723   724   725   729   731   732   733
##    74    25     5    29     7     8    26    26    23    83
##   734   735   739   741   742   743   744   745   749   751
##    34    13    62    59    27     5     1    17    16     3
##   752   753   754   755   756   759   760   761   769   791
##    24     3     1     8    10    74  1000     6     3   175
##   792   793   794   799   801   802   803   811   812   821
##   142   146     5   966    21     2     3     5   322     7
##   822   823   831   832   833   834   835   836   837   838
##    55     1     9    23    18    29    34    45   407    84
##   839   841   842   843   844   845   846   847   848   849
##    27    21   103   104    38    44  1993   426  2114    18
##   850   852   853   854   855   856   860   861   862   863
##   356    16    22     2   201    18   860  5750   489   344
##   864   866   867   868   869   870   871   872   890   891
##    16    64     2    21    13    11   111    41    46  2613
##   893   896   897   898   899  899A   901   902   903   904
##   148 87632   128  3535     1     5 12177  1034   394   214
##   905   906   907   909   911   912   914   921   922   923
##   299  3579    69  1651   170   453    76    87    29     1
##   924   931   932   941   942   943   944   945   946   994
##     5     1     4    82    21     3    13    11   222  1293
##   999  <NA>
```

```
##    1217   8488
# INCIDENT_NUMBER & FD_STATION have a lot of similarities...
sum(I$INCIDENT_NUMBER == substr(I$FD_STATION,0,9), na.rm=T) #375456
```

```
## [1] 375456
```

```
length(unique(I$FD_STATION)) #375478
```

```
## [1] 375478
```

```
# https://www.quora.com/How-do-I-get-a-frequency-count-based-on-two-columns-variables-in-an-R-dataframe

# I'm trying to build a cross-table

I$PROPERTY_GROUP <- vector(mode='character', length=nrow(I))
# I[I$PROPERTY]
PROPERTY_GROUP <- list()
PROPERTY_GROUP$A <- (101:199)
PROPERTY_GROUP$B <- (201:299)
PROPERTY_GROUP$C <- (301:399)
PROPERTY_GROUP$D <- (401:499)
PROPERTY_GROUP$E <- (501:599)
PROPERTY_GROUP$F <- (601:799)
PROPERTY_GROUP$O <- (801:999)

for(g in c(LETTERS[(1:6)],'O')){
  I$PROPERTY_GROUP[I$PROPERTY %in% PROPERTY_GROUP[[g]]] <- g
}

table(format(I$INCIDENT_DATE, '%A'), I$PROPERTY_GROUP) # This is Day of week vs Property Group
```

```
##
##                   A      B      C      D      E      F      O
##    Friday      1456   8274   7828  58514   4169   3211   1780  20139
##    Monday      1407   7413   7878  59931   4370   2703   1788  17403
##    Saturday    1430   7038   7249  61043   2737   3215   1377  20011
##    Sunday      1425   6604   7305  61913   2016   2442   1107  19042
##    Thursday    1404   8142   7891  56837   4530   2980   1808  18761
##    Tuesday     1356   7587   7636  58239   4646   3007   1787  17844
##    Wednesday   1390   7890   7749  57293   4447   2852   1778  18298
```

```
MEDICAL <- I[I$EVENT_TYPE %in% 'Medical',]
table(format(MEDICAL$INCIDENT_DATE, '%A'), MEDICAL$PROPERTY_GROUP)
```

```
##
##                  A      B      C      D      E      F      O
##    Friday      935   3990   5872  30795   1780   1595   377   6466
##    Monday      921   3324   5902  31551   1852   1281   365   5278
##    Saturday    907   3589   5454  31600    888   1704   190   7573
##    Sunday      908   3398   5576  32531    558   1163   137   7018
##    Thursday    887   3876   5962  30179   2137   1461   429   5760
##    Tuesday     865   3407   5633  30362   2095   1423   343   5190
##    Wednesday   901   3654   5692  29876   1916   1413   383   5420
```

```
# This below function calculates time difference between time call was received
# and recorded onscene time. I accounted for cases where the date would roll over between call & arriva
```

```r
# However, I later realized that the data itself is encoded to account for this:
# DISPATCH_DATE includes the date & time of ONSCENE whereas INCIDENT_DATE is the date of call.

# tta <- function(x){
#   # print(x);
#     # print(x$INITIAL_CALL_HOUR)
#     if(!is.na(x$INITIAL_CALL_HOUR) & !is.na(x$ONSCENE_HOUR)){
#       # if(x$ONSCENE_HOUR < x$INITIAL_CALL_HOUR){
#         # ic_time <- paste(x$INITIAL_CALL_HOUR, x$INITIAL_CALL_MIN, x$INITIAL_CALL_SEC, sep=" ")
#         ic_time <- paste(as.Date(x$INCIDENT_DATE), " ",formatC(x$INITIAL_CALL_HOUR, width=2, flag="0"),
#         os_time <- paste(as.Date(x$DISPATCH_DATE)," ",formatC(x$ONSCENE_HOUR, width=2, flag="0"), ":", 
#         # os <- as.POSIXct()
#         # cat(os_time, ic_time)
#         return(as.POSIXct(os_time, format="%Y-%m-%d %H:%M:%S") - as.POSIXct(ic_time, format="%Y-%m-%d %
#     # }
#
#
#   # }else{
#
#   # }else{
#
#   }
# }

# This function works if you pass it a single row from the data.frame, such as I[1695,]
# However, there are problems if you try to run it against the entire dataset.
# tta <- function(x){
#   onscene_date <- x$INCIDENT_DATE
#   if(x$ONSCENE_HOUR < x$INITIAL_CALL_HOUR){ onscene_date <- as.Date(x$INCIDENT_DATE) + 1 }
#   ic_time <- paste(as.Date(x$INCIDENT_DATE), " ",formatC(x$INITIAL_CALL_HOUR, width=2, flag="0"), ":"
#   os_time <- paste(as.Date(onscene_date)," ",formatC(x$ONSCENE_HOUR, width=2, flag="0"), ":", formatC
#   # cat(ic_time,"\n",os_time,"\n")
#   return(as.POSIXct(os_time, format="%Y-%m-%d %H:%M:%S") - as.POSIXct(ic_time, format="%Y-%m-%d %H:%M
# }

tta <- function(x){

  if(sum(is.na(c(x["INCIDENT_DATE"], x["INITIAL_CALL_HOUR"], x["INITIAL_CALL_MIN"], x["INITIAL_CALL_SEC
  onscene_date <- as.Date(x["INCIDENT_DATE"])
  if(x["ONSCENE_HOUR"] < x["INITIAL_CALL_HOUR"]){ onscene_date <- as.Date(x["INCIDENT_DATE"]) + 1 }
  ic_time <- paste(as.Date(x["INCIDENT_DATE"]), " ",formatC(x["INITIAL_CALL_HOUR"], width=2, flag="0"),
  os_time <- paste(as.Date(onscene_date)," ",formatC(x["ONSCENE_HOUR"], width=2, flag="0"), ":", formatC
  # cat(ic_time,"\n",os_time,"\n")
  return(as.POSIXct(os_time, format="%Y-%m-%d %H:%M:%S") - as.POSIXct(ic_time, format="%Y-%m-%d %H:%M:%S
}
# I[!is.na(I$INITIAL_CALL_HOUR) & !is.na(I$ONSCENE_HOUR) & I$ONSCENE_HOUR < I$INITIAL_CALL_HOUR,] #Roll

# I$TTA <- apply(I, 1, tta) # This is how I obtained the TTA info... runtime: about 20 minutes
TTA <- read.csv(file="../csv/TTA.csv", header = T)
I <- cbind(I, TTA$TTA)


# Trying to test the above: Boolean "is DISPATCH_DATE" time always ONSCENE time?
```

```r
testColumn <- function(x){
  os_time <- paste(as.Date(x$DISPATCH_DATE)," ",formatC(x$ONSCENE_HOUR, width=2, flag="0"), ":", formatC
  return(as.POSIXct(os_time, format="%Y-%m-%d %H:%M:%S") == x$DISPATCH_DATE)
  # cat(x$DISPATCH_DATE)
}

sum(testColumn(I[1:nrow(I),]), na.rm=T) #563070
```

```
## [1] 563070
```

```r
sum(is.na(testColumn(I[1:nrow(I),]))) #17275
```

```
## [1] 17275
```

```r
sum(I$DISPATCH_DATE > I$INCIDENT_DATE, na.rm=T) # 703568
```

```
## [1] 703568
```

```r
sum(is.na(I$DISPATCH_DATE > I$INCIDENT_DATE)) # 16802
```

```
## [1] 16802
```

```r
aggregate(I$TTA, list(I$PROPERTY_GROUP), mean, na.rm=T)
```

```
##   Group.1        x
## 1          6.407133
## 2       A 5.133175
## 3       B 4.990055
## 4       C 5.429354
## 5       D 5.300845
## 6       E 5.164147
## 7       F 5.819520
## 8       O 6.278032
```

```r
table(I$PROPERTY_GROUP)
```

```
##
##          A      B      C      D      E      F      O
##   9868  52948  53536 413770  26915  20410  11425 131498
```

```r
summary(I)
```

```
##  INCIDENT_NUMBER                      EVENT_TYPE
##  Length:720370     Medical                      :354742
##  Class :character  Alarm Highrise Residential   : 46464
##  Mode  :character  Alarm Residential            : 28734
##                    Alarm Commercial/Industrial  : 25569
##                    Vehicle - Personal Injury    : 24850
##                    Carbon Monoxide - Non Medical: 20502
##                    (Other)                      :219509
##  DISPATCH_DATE                 EVENT_TYPE_CD              MAIN_STREET
##  Min.   :2011-01-01 00:09:02   Medical:158831                    :323855
##  1st Qu.:2012-03-30 21:09:21   MEU    : 64176   NA               : 96701
##  Median :2013-11-22 20:04:02   METB   : 60585   YONGE ST         :  5798
##  Mean   :2013-11-19 17:23:44   FAHR   : 46474   LAWRENCE AVE E:  3815
##  3rd Qu.:2015-06-10 18:44:29   FAR    : 28735   BATHURST ST    :  3361
##  Max.   :2017-01-01 00:04:56   FACI   : 25573   SHERBOURNE ST :  2996
##  NA's   :16802                  (Other):335996   (Other)        :283844
```

```
##       CROSS_STREET          FSA          RESPONDING_UNITS
##            :300076            :296120  Min.   :  1.000
## NA         : 90273  NA        : 59018  1st Qu.:  1.000
## YONGE ST   :  6409  M5A       : 10398  Median :  1.000
## BLOOR ST W :  2343  M6K       :  7653  Mean   :  2.283
## BATHURST ST:  2270  M1E       :  6976  3rd Qu.:  4.000
## DUNDAS ST E:  2188  M1P       :  6714  Max.   :453.000
## (Other)    :316811  (Other)   :333491  NA's   :32
##   FD_STATION       OFM_INVESTIGATIONS_CONTACTED AID_TO_FROM_OTHER_DEPTS
## Length:720370     Min.   :0.000000             Min.   :1.000
## Class :character  1st Qu.:0.000000             1st Qu.:4.000
## Mode  :character  Median :0.000000             Median :4.000
##                   Mean   :0.001588             Mean   :3.948
##                   3rd Qu.:0.000000             3rd Qu.:4.000
##                   Max.   :1.000000             Max.   :4.000
##
##   INCIDENT_DATE                   INITIAL_CALL_HOUR INITIAL_CALL_MIN
## Min.   :2011-01-01 00:00:00   Min.   : 0.00    Min.   : 0.00
## 1st Qu.:2012-04-02 00:00:00   1st Qu.: 9.00    1st Qu.:14.00
## Median :2013-11-27 00:00:00   Median :13.00    Median :29.00
## Mean   :2013-11-22 06:26:32   Mean   :12.95    Mean   :29.49
## 3rd Qu.:2015-06-14 00:00:00   3rd Qu.:18.00    3rd Qu.:44.00
## Max.   :2016-12-31 00:00:00   Max.   :23.00    Max.   :59.00
##
##  INITIAL_CALL_SEC DISPATCH_HOUR    DISPATCH_MIN    DISPATCH_SEC
## Min.   : 0.00    Min.   : 0.00   Min.   : 0.0   Min.   : 0.0
## 1st Qu.:14.00    1st Qu.: 9.00   1st Qu.:15.0   1st Qu.:15.0
## Median :30.00    Median :13.00   Median :30.0   Median :29.0
## Mean   :29.49    Mean   :12.95   Mean   :29.5   Mean   :29.5
## 3rd Qu.:45.00    3rd Qu.:18.00   3rd Qu.:44.0   3rd Qu.:45.0
## Max.   :59.00    Max.   :23.00   Max.   :59.0   Max.   :59.0
##                  NA's   :177     NA's   :177    NA's   :177
##   ONSCENE_HOUR    ONSCENE_MIN     ONSCENE_SEC    INITIAL_UNIT_PERSONNEL
## Min.   : 0.00   Min.   : 0.00   Min.   : 0.00   Min.   : 0.000
## 1st Qu.: 9.00   1st Qu.:15.00   1st Qu.:14.00   1st Qu.: 4.000
## Median :14.00   Median :30.00   Median :29.00   Median : 4.000
## Mean   :12.97   Mean   :29.53   Mean   :29.45   Mean   : 3.824
## 3rd Qu.:18.00   3rd Qu.:45.00   3rd Qu.:44.00   3rd Qu.: 4.000
## Max.   :23.00   Max.   :59.00   Max.   :59.00   Max.   :75.000
## NA's   :15371   NA's   :15371   NA's   :15371
## TOTAL_NUM_PERSONNEL     EST_KM         ALARM_TO_FD       RESPONSE_TYPE
## Min.   :   0.000   Min.   :-1.000   3      :397304   89     :336333
## 1st Qu.:   4.000   1st Qu.: 2.000   1      :148280   32     : 42456
## Median :   4.000   Median : 2.000   5      :126343   31     : 39683
## Mean   :   8.043   Mean   : 2.537   4      : 20673   62     : 34384
## 3rd Qu.:  13.000   3rd Qu.: 3.000   2      : 16628   34     : 27929
## Max.   :1277.000   Max.   :99.000   7      :  3925   35     : 18733
##                                     (Other):  7217   (Other):220852
##     RESCUES           FF_INJURIES      FF_FATALITIES AGENT_APP_HOUR
## Min.   : 0.00000   Min.   :0.000000   Min.   :0     Min.   : 0
## 1st Qu.: 0.00000   1st Qu.:0.000000   1st Qu.:0     1st Qu.: 8
## Median : 0.00000   Median :0.000000   Median :0     Median :14
## Mean   : 0.02654   Mean   :0.000373   Mean   :0     Mean   :13
## 3rd Qu.: 0.00000   3rd Qu.:0.000000   3rd Qu.:0     3rd Qu.:19
```

```
## Max.   :43.00000  Max.   :4.000000  Max.    :0       Max.    :23
##                                                      NA's   :708040
## AGENT_APP_MIN   AGENT_APP_SEC    CONTROL_DATE      CONTROL_HOUR
## Min.   : 0.0    Min.   : 0.0   Length:720370     Min.   : 0
## 1st Qu.:14.0    1st Qu.: 0.0   Class :character  1st Qu.: 8
## Median :29.5    Median : 4.0   Mode  :character  Median :14
## Mean   :29.2    Mean   :15.9                     Mean   :13
## 3rd Qu.:44.8    3rd Qu.:30.0                      3rd Qu.:19
## Max.   :59.0    Max.   :59.0                      Max.   :23
## NA's   :708040  NA's   :708040                    NA's   :707821
##  CONTROL_MIN     CONTROL_SEC    STATUS_ON_ARRIVAL WATER
## Min.   : 0      Min.   : 0.0   NA     :708040    1 : 8190
## 1st Qu.:14      1st Qu.: 0.0   1      :  3244    2 :   462
## Median :30      Median :13.0   3      :  3213    3 :   128
## Mean   :29      Mean   :18.8   2      :  2322    4 :    81
## 3rd Qu.:45      3rd Qu.:35.0   4      :  1785    5 :  2638
## Max.   :59      Max.   :59.0   7      :  1167    9 :   831
## NA's   :707821  NA's   :707821 (Other):   599    NA:708040
## FIRE_CONTROL     PROPERTY      AREA_OF_ORIGIN    IGNITION_SOURCE
## 1 :  8612     323   :204952   NA     :708040   NA     :708040
## 2 :   373     301   :122844   24     :  2110   999    :  3801
## 3 :  2422     896   : 87632   81     :  1327   11     :  1394
## 4 :   594     302   : 26743   44     :   913   71     :  1216
## 5 :   329     223   : 23119   64     :   716   81     :   742
## NA:708040     303   : 21600   22     :   665   82     :   413
##               (Other):233480   (Other):  6599   (Other):  4764
## FUEL_OF_IGNITION_SOURCE OBJECT_OR_MATERIAL_FIRST_IGNITED POSSIBLE_CAUSE
## NA     :708040          NA     :708040                 NA     :708040
## 99     :  3979          99     :  1838                 99     :  2219
## 4      :  2851          46     :  1612                 52     :  1780
## 97     :  1326          74     :  1071                 45     :  1310
## 13     :  1278          47     :   840                 44     :  1164
## 7      :   858          43     :   790                 1      :   904
## (Other):  2038          (Other):  6179                 (Other):  4953
##  VEH_PURPOSE      VEH_FUEL       EST_LOSS         INSURANCE_ESTIMATE
## NA     :718631  NA     :718631  Min.   :       0  1 :   215
## 1      :  1311  1      :  1376  1st Qu.:       0  2 :  5594
## 6      :   149  2      :   262  Median :       0  8 :  2350
## 7      :   122  99     :    35  Mean   :     465  9 :  4253
## 9      :    95  4      :    31  3rd Qu.:       0  NA:707958
## 99     :    33  97     :    15  Max.   :13000000
## (Other):    29  (Other):    20
## EST_VALUE_AT_RISK CIVILIAN_FIRE_INJURY CIVILIAN_FIRE_FATALITY
## NA     :707958    Min.   : 0.000000   Min.   :0.00e+00
## 2      :  5292    1st Qu.: 0.000000   1st Qu.:0.00e+00
## 3      :  1821    Median : 0.000000   Median :0.00e+00
## 1      :  1581    Mean   : 0.001224   Mean   :8.47e-05
## 99     :  1266    3rd Qu.: 0.000000   3rd Qu.:0.00e+00
## 6      :   603    Max.   :12.000000   Max.   :3.00e+00
## (Other):  1849
## RESCUED_CHILDREN  RESCUED_ADULTS    RESCUED_SENIORS
## Min.   :0.000000  Min.   :0.0e+00  Min.   :0.00e+00
## 1st Qu.:0.000000  1st Qu.:0.0e+00  1st Qu.:0.00e+00
## Median :0.000000  Median :0.0e+00  Median :0.00e+00
```

```
## Mean   :0.000136   Mean   :9.2e-04   Mean   :2.04e-04
## 3rd Qu.:0.000000   3rd Qu.:0.0e+00   3rd Qu.:0.00e+00
## Max.   :6.000000   Max.   :4.0e+01   Max.   :1.00e+01
##
## PHYSICAL_CONDITION_1 PHYSICAL_CONDITION_2 PHYSICAL_CONDITION_3
## NA     :708041      1 :    14            1 :     2
## 98     :  6795      3 :     1            2 :     1
## 1      :  5073      4 :     1            99:     4
## 99     :   300      7 :     3            NA:720363
## 3      :    61      98:    22
## 7      :    43      99:     6
## (Other):    57      NA:720323
## CIV_FIRE_CONTROL CIV_EVACUATION CIV_EVACUATION_REASON_1
## 1 :  3677         1 :  4672      NA     :708041
## 2 :   997         2 :  1210      8      :  9246
## 3 :  4136         3 :  2494      1      :  1570
## 4 :  2590         4 :  3598      9      :  1263
## 9 :   929         9 :   355      6      :    81
## NA:708041         NA:708041      4      :    59
##                                  (Other):   110
## CIV_EVACUATION_REASON_2 OPP        MOE          TSSA         ESA
## NA     :720265          1 :   939  1 :  1359    1 :   548    1 :   290
## 9      :    42          NA:719431  NA:719011    NA:719822    NA:720080
## 1      :    18
## 8      :    18
## 3      :    11
## 4      :    11
## (Other):     5
## MOL         EMS          CANUTEC      GAS          HYDRO
## 1 :   290   1 :    45    1 :   232    1 :  5220    1 :  4404
## NA:720080   NA:720325    NA:720138    NA:715150    NA:715966
##
##
##
##
##
## MUNICIPAL_BUILDING_OFFICE MUNICIPAL_HEALTH_OFFICE MUNICIPAL_POLICE
## 1 :   228                 1 :  1362               1 :  1397
## NA:720142                 NA:719008               NA:718973
##
##
##
##
##
## OTHER        INITIAL_DETECTION  EXTENT_FIRE        EXTENT_SMOKE
## 1 :  2982    NA     :711584     NA     :711585     NA     :711585
## NA:717388    6      :  5672     1      :  4135     2      :  2342
##             1      :  1503     2      :  3107     4      :  2126
##             2      :   874     3      :   374     7      :  1414
##             3      :   372     4      :   336     3      :   686
##             9      :   214     9      :   273     8      :   564
##             (Other):   151     (Other):   560     (Other):  1653
## EST_NUM_PERSONS_DISPLACED POSSIBLE_BUSINESS_IMPACT    COMPLEX
## Min.   :  0.0000          NA     :711585             Min.   : 1.0
```

```
##   1st Qu.:  0.0000              8    :  3905         1st Qu.: 3.0
##   Median :  0.0000              1    :  3662         Median :20.0
##   Mean   :  0.2167             9    :   569         Mean   :50.8
##   3rd Qu.:  0.0000             2    :   485         3rd Qu.:98.0
##   Max.   :999.0000             3    :    91         Max.   :99.0
##                            (Other):    73           NA's   :711589
##    OCC_STATUS      OCC_TYPE       BLD_STATUS       BLD_HEIGHT
##   NA     :711589  1 :  3162   NA     :711589   0     :711589
##   1      :  6835  2 :  3824   1      :  7656   2     :  2359
##   8      :   873  3 :    90   8      :   514   1     :  1637
##   2      :   853  4 :   577   2      :   362   3     :  1120
##   9      :   123  8 :   794   3      :   117   999   :   421
##   3      :    48  9 :   334   9      :    61   4     :   322
##   (Other):    49  NA:711589   (Other):    71   (Other):  2922
##   LEVEL_OF_ORIGIN  AGE_OF_STRUCTURE
##          :567697   NA     :711589
##   NA     :143892   4      :  3071
##   001    :  3437   3      :  2466
##   B01    :  1226   9      :  1522
##   002    :  1095   2      :   900
##   003    :   503   5      :   535
##   (Other):  2520   (Other):   287
##   SMOKE_ALARM_PRESENCE_AND_OPERATION_MAIN_FLOOR
##   1 :  1980
##   2 :  3763
##   3 :  1373
##   4 :   574
##   9 :  1091
##   NA:711589
##
##   SMOKE_ALARM_FAILURE_TO_OPERATE SMOKE_ALARM_TYPE
##   NA     :711589                1 :  1781
##   98     :  5991                2 :  1786
##   4      :   925                3 :    18
##   99     :   501                4 :  1663
##   9      :   482                8 :  2080
##   5      :   318                9 :  1453
##   (Other):   564                NA:711589
##   SMOKE_ALARM_OTHER_FLOOR_PRESENCE SMOKE_ALARM_ON_ALL_FLOORS
##   1 :   535                       1 :  4145
##   2 :  2226                       2 :   356
##   3 :  1956                       3 :   348
##   4 :   768                       4 :   375
##   8 :  1519                       8 :  1225
##   9 :  1777                       9 :  2332
##   NA:711589                       NA:711589
##   SMOKE_ALARM_IMPACT_ON_EVAC SMOKE_ALARM_IMPACT_ON_NUM_EVAC
##   NA     :711589             Min.   : 0.0000
##   1      :  2316             1st Qu.: 0.0000
##   7      :  1882             Median : 0.0000
##   8      :  1805             Mean   : 0.2307
##   9      :  1105             3rd Qu.: 0.0000
##   2      :   897             Max.   :99.0000
##   (Other):   776
```

```
##  FIRE_ALARM_SYSTEM_PRESENCE FIRE_ALARM_SYSTEM_OPERATION
##  1 :  4903                   1 :   3557
##  2 :  1388                   2 :   1098
##  8 :  1498                   8 :   2988
##  9 :   992                   9 :   1138
##  NA:711589                   NA:711589
##
##
##  FIRE_ALARM_SYSTEM_IMPACT SPRINKLER_SYSTEM_PRESENCE
##  NA     :711589           1 :   1331
##  8      :  2441           2 :    907
##  7      :  1923           3 :   5104
##  1      :  1505           9 :   1439
##  9      :  1301           NA:711589
##  2      :   999
##  (Other):   612
##  SPRINKLER_SYSTEM_ACTIVATION    ROW_NAS        PROPERTY_GROUP
##  NA     :711589             Min.   :12.00   Length:720370
##  8      :  5393             1st Qu.:58.00   Class :character
##  2      :  1294             Median :58.00   Mode  :character
##  1      :   667             Mean   :58.22
##  9      :   664             3rd Qu.:58.00
##  3      :   591             Max.   :70.00
##  (Other):   172
##     TTA$TTA
##  Min.   : 0.000
##  1st Qu.: 4.133
##  Median : 5.050
##  Mean   : 5.524
##  3rd Qu.: 6.150
##  Max.   :59.967
##  NA's   :15374
```