r_script.R

jingxuefeng

Mon May 8 15:51:43 2017

```
library(foreign)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(ggplot2)
library(ggmap)
library(shiny)
library(data.table)
```

```
## ------------------------------------------------------------------------------
```

```
## data.table + dplyr code now lives in dtplyr.
## Please library(dtplyr)!
```

```
## ------------------------------------------------------------------------------
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
#After read through "Data ANalysis Guidelines and Issues", I find out that fireincident and basicincide
#read tables that has relation with "residential fires with fatalities"

incident_07 <- read.dbf("NFIRS_2007/fireincident.dbf")
basic_07 <- read.dbf("NFIRS_2007/basicincident.dbf")
incident_08 <- read.dbf("NFIRS_2008/fireincident.dbf")
basic_08 <- read.dbf("NFIRS_2008/basicincident.dbf")
incident_09 <- read.dbf("NFIRS_2009/fireincident.dbf")
basic_09 <- read.dbf("NFIRS_2009/basicincident.dbf")
incident_10 <- read.dbf("NFIRS_2010/fireincident.dbf")
basic_10 <- read.dbf("NFIRS_2010/basicincident.dbf")
incident_11 <- read.dbf("NFIRS_2011/fireincident.dbf")
basic_11 <- read.dbf("NFIRS_2011/basicincident.dbf")
incident_12 <- fread( "NFIRS_2012/fireincident.txt", sep = "^")
```

```
##
Read 0.0% of 599879 rows
Read 36.7% of 599879 rows
```

```
Read 61.7% of 599879 rows
Read 599879 rows and 80 (of 80) columns from 0.145 GB file in 00:00:05
```

```
basic_12 <- fread( "NFIRS_2012/basicincident.txt", sep = "^")
```

```
##
Read 0.0% of 2120288 rows
Read 14.6% of 2120288 rows
Read 30.2% of 2120288 rows
Read 45.7% of 2120288 rows
Read 60.8% of 2120288 rows
Read 75.0% of 2120288 rows
Read 88.7% of 2120288 rows
Read 2120288 rows and 41 (of 41) columns from 0.350 GB file in 00:00:09
```

```
incident_13 <- fread( "NFIRS_2013/fireincident.txt", sep = "^")
```

```
##
Read 46.9% of 554671 rows
Read 554671 rows and 80 (of 80) columns from 0.135 GB file in 00:00:03
```

```
basic_13 <- fread( "NFIRS_2013/basicincident.txt", sep = "^")
```

```
##
Read 0.0% of 2003907 rows
Read 14.0% of 2003907 rows
Read 27.9% of 2003907 rows
Read 42.9% of 2003907 rows
Read 57.4% of 2003907 rows
Read 70.9% of 2003907 rows
Read 84.3% of 2003907 rows
Read 97.8% of 2003907 rows
Read 2003907 rows and 41 (of 41) columns from 0.331 GB file in 00:00:13
```

```
incident_14 <- fread( "NFIRS_2014/fireincident.txt", sep = "^")
```

```
##
Read 48.6% of 596521 rows
Read 98.9% of 596521 rows
Read 596521 rows and 80 (of 80) columns from 0.144 GB file in 00:00:04
```

```
basic_14 <- fread( "NFIRS_2014/basicincident.txt", sep = "^")
```

```
##
Read 0.0% of 2116746 rows
Read 9.4% of 2116746 rows
Read 20.8% of 2116746 rows
Read 33.1% of 2116746 rows
Read 45.4% of 2116746 rows
Read 57.6% of 2116746 rows
Read 69.9% of 2116746 rows
Read 81.7% of 2116746 rows
Read 93.1% of 2116746 rows
Read 2116746 rows and 41 (of 41) columns from 0.349 GB file in 00:00:12
```

```
#from "Public Data Release Format" document I leanred that civilian death is severity = 5 in Civilian C
#read civiliancausitytable
civilian_07 <- read.dbf("NFIRS_2007/civiliancasualty.dbf")
```

```
#TABLES
#Clean Data

#since column "Not_RES" shows whether the fire incident is in resident or not
#delete data that are not in residential

incident_07 = incident_07[incident_07$NOT_RES == "N",]
incident_08 = incident_08[incident_08$NOT_RES == "N",]
incident_09 = incident_09[incident_09$NOT_RES == "N",]
incident_10 = incident_10[incident_10$NOT_RES == "N",]
incident_11 = incident_11[incident_11$NOT_RES == "N",]
incident_12 = incident_12[incident_12$NOT_RES == "N",]
incident_13 = incident_13[incident_13$NOT_RES == "N",]
incident_14 = incident_14[incident_14$NOT_RES == "N",]




#remove those data that is not fatality
basic_07 = basic_07[basic_07$FF_DEATH != 0 | basic_07$OTH_DEATH != 0, ]
basic_08 = basic_08[basic_08$FF_DEATH != 0 | basic_08$OTH_DEATH != 0, ]
basic_09 = basic_09[basic_09$FF_DEATH != 0 | basic_09$OTH_DEATH != 0, ]
basic_10 = basic_10[basic_11$FF_DEATH != 0 | basic_10$OTH_DEATH != 0, ]
```

```
## Warning in basic_11$FF_DEATH != 0 | basic_10$OTH_DEATH != 0: longer object
## length is not a multiple of shorter object length
```

```
basic_11 = basic_11[basic_11$FF_DEATH != 0 | basic_11$OTH_DEATH != 0, ]
basic_12 = basic_12[basic_12$FF_DEATH != 0 | basic_12$OTH_DEATH != 0, ]
basic_13 = basic_13[basic_13$FF_DEATH != 0 | basic_13$OTH_DEATH != 0, ]
basic_14 = basic_14[basic_14$FF_DEATH != 0 | basic_14$OTH_DEATH != 0, ]




#combine those tables
fatalres_07 <- inner_join(basic_07, incident_07)
```

```
## Joining, by = c("STATE", "FDID", "INC_DATE", "INC_NO", "EXP_NO", "VERSION")
```

```
## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
fatalres_08 <- inner_join(basic_08, incident_08)
```

```
## Joining, by = c("STATE", "FDID", "INC_DATE", "INC_NO", "EXP_NO", "VERSION")
```

```
## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
```

```
## factors with different levels, coercing to character vector

## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
fatalres_09 <- inner_join(basic_09, incident_09)

## Joining, by = c("STATE", "FDID", "INC_DATE", "INC_NO", "EXP_NO", "VERSION")

## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector

## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
fatalres_10 <- inner_join(basic_10, incident_10)

## Joining, by = c("STATE", "FDID", "INC_DATE", "INC_NO", "EXP_NO", "VERSION")

## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector

## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
fatalres_11 <- inner_join(basic_11, incident_11)

## Joining, by = c("STATE", "FDID", "INC_DATE", "INC_NO", "EXP_NO", "VERSION")

## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector

## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
fatalres_12 <- inner_join(basic_12, incident_12)

## Joining, by = c("STATE", "FDID", "INC_DATE", "INC_NO", "EXP_NO", "VERSION")
fatalres_13 <- inner_join(basic_13, incident_13)

## Joining, by = c("STATE", "FDID", "INC_DATE", "INC_NO", "EXP_NO", "VERSION")
fatalres_14 <- inner_join(basic_14, incident_14)

## Joining, by = c("STATE", "FDID", "INC_DATE", "INC_NO", "EXP_NO", "VERSION")
## Since I want to focus on cause of residential fire with fatalities,
## I think "STATE', "FDID", "INC_DATE", "AREA_ORIG", "HEAT_SOURC", "FIRST_IGN", "CAUSE_IGN"


colnames(fatalres_07)

##   [1] "STATE"      "FDID"       "INC_DATE"   "INC_NO"     "EXP_NO"
##   [6] "VERSION"    "DEPT_STA"   "INC_TYPE"   "ADD_WILD"   "AID"
##  [11] "ALARM"      "ARRIVAL"    "INC_CONT"   "LU_CLEAR"   "SHIFT"
##  [16] "ALARMS"     "DISTRICT"   "ACT_TAK1"   "ACT_TAK2"   "ACT_TAK3"
##  [21] "APP_MOD"    "SUP_APP"    "EMS_APP"    "OTH_APP"    "SUP_PER"
##  [26] "EMS_PER"    "OTH_PER"    "RESOU_AID"  "PROP_LOSS"  "CONT_LOSS"
##  [31] "PROP_VAL"   "CONT_VAL"   "FF_DEATH"   "OTH_DEATH"  "FF_INJ"
##  [36] "OTH_INJ"    "DET_ALERT"  "HAZ_REL"    "MIXED_USE"  "PROP_USE"
```

```
##  [41] "CENSUS"     "NUM_UNIT"   "NOT_RES"    "BLDG_INVOL" "ACRES_BURN"
##  [46] "LESS_1ACRE" "ON_SITE_M1" "MAT_STOR1"  "ON_SITE_M2" "MAT_STOR2"
##  [51] "ON_SITE_M3" "MAT_STOR3"  "AREA_ORIG"  "HEAT_SOURC" "FIRST_IGN"
##  [56] "CONF_ORIG"  "TYPE_MAT"   "CAUSE_IGN"  "FACT_IGN_1" "FACT_IGN_2"
##  [61] "HUM_FAC_1"  "HUM_FAC_2"  "HUM_FAC_3"  "HUM_FAC_4"  "HUM_FAC_5"
##  [66] "HUM_FAC_6"  "HUM_FAC_7"  "HUM_FAC_8"  "AGE"        "SEX"
##  [71] "EQUIP_INV"  "SUP_FAC_1"  "SUP_FAC_2"  "SUP_FAC_3"  "MOB_INVOL"
##  [76] "MOB_TYPE"   "MOB_MAKE"   "MOB_MODEL"  "MOB_YEAR"   "MOB_LIC_PL"
##  [81] "MOB_STATE"  "MOB_VIN_NO" "EQ_BRAND"   "EQ_MODEL"   "EQ_SER_NO"
##  [86] "EQ_YEAR"    "EQ_POWER"   "EQ_PORT"    "FIRE_SPRD"  "STRUC_TYPE"
##  [91] "STRUC_STAT" "BLDG_ABOVE" "BLDG_BELOW" "BLDG_LGTH"  "BLDG_WIDTH"
##  [96] "TOT_SQ_FT"  "FIRE_ORIG"  "ST_DAM_MIN" "ST_DAM_SIG" "ST_DAM_HVY"
## [101] "ST_DAM_XTR" "FLAME_SPRD" "ITEM_SPRD"  "MAT_SPRD"   "DETECTOR"
## [106] "DET_TYPE"   "DET_POWER"  "DET_OPERAT" "DET_EFFECT" "DET_FAIL"
## [111] "AES_PRES"   "AES_TYPE"   "AES_OPER"   "NO_SPR_OP"  "AES_FAIL"
```

```
#keep column 1~4, 53, 55, 58

fatalres_07 <- fatalres_07[c(1:4,53:55,58)]
fatalres_08 <- fatalres_08[c(1:4,53:55,58)]
fatalres_09 <- fatalres_09[c(1:4,53:55,58)]
fatalres_10 <- fatalres_10[c(1:4,53:55,58)]
fatalres_11 <- fatalres_11[c(1:4,53:55,58)]
fatalres_12 <- fatalres_12[c(1:4,53:55,58)]
fatalres_13 <- fatalres_13[c(1:4,53:55,58)]
fatalres_14 <- fatalres_14[c(1:4,53:55,58)]



#Add specific address to the table
#read address tables
address_07 <- read.dbf("NFIRS_2007/incidentaddress.dbf")
address_08 <- read.dbf("NFIRS_2008/incidentaddress.dbf")
address_09 <- read.dbf("NFIRS_2009/incidentaddress.dbf")
address_10 <- read.dbf("NFIRS_2010/incidentaddress.dbf")
address_11 <- read.dbf("NFIRS_2011/incidentaddress.dbf")
address_12 <- fread( "NFIRS_2012/incidentaddress.txt", sep = "^")
```

```
##
Read 9.9% of 2120288 rows
Read 26.4% of 2120288 rows
Read 42.4% of 2120288 rows
Read 59.0% of 2120288 rows
Read 74.5% of 2120288 rows
Read 97.2% of 2120288 rows
Read 100.0% of 2120288 rows
Read 2120288 rows and 17 (of 17) columns from 0.204 GB file in 00:00:09
```

```
address_13 <- read.table( "NFIRS_2013/incidentaddress.txt", sep = "^", header = TRUE)
```

```
## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec =
## dec, : embedded nul(s) found in input
```

```
address_14 <- fread( "NFIRS_2014/incidentaddress.txt", sep = "^")
```

```
##
```

```
Read 2.8% of 2116746 rows
Read 11.3% of 2116746 rows
Read 25.5% of 2116746 rows
Read 42.0% of 2116746 rows
Read 58.1% of 2116746 rows
Read 73.2% of 2116746 rows
Read 75.1% of 2116746 rows
Read 95.0% of 2116746 rows
Read 2116746 rows and 17 (of 17) columns from 0.204 GB file in 00:00:13
```

```
#combine address and previous table

fatalres_07 <- inner_join(fatalres_07, address_07)
```

```
## Joining, by = c("STATE", "FDID", "INC_DATE", "INC_NO")
```

```
## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## character vector and factor, coercing into character vector
```

```
## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## character vector and factor, coercing into character vector
```

```
fatalres_08 <- inner_join(fatalres_08, address_08)
```

```
## Joining, by = c("STATE", "FDID", "INC_DATE", "INC_NO")
```

```
## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## character vector and factor, coercing into character vector
```

```
## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## character vector and factor, coercing into character vector
```

```
fatalres_09 <- inner_join(fatalres_09, address_09)
```

```
## Joining, by = c("STATE", "FDID", "INC_DATE", "INC_NO")
```

```
## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## character vector and factor, coercing into character vector
```

```
## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## character vector and factor, coercing into character vector
```

```
fatalres_10 <- inner_join(fatalres_10, address_10)
```

```
## Joining, by = c("STATE", "FDID", "INC_DATE", "INC_NO")
```

```
## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## character vector and factor, coercing into character vector
```

```
## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## character vector and factor, coercing into character vector
```

```
fatalres_11 <- inner_join(fatalres_11, address_11)
```

```
## Joining, by = c("STATE", "FDID", "INC_DATE", "INC_NO")
```

```
## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## character vector and factor, coercing into character vector
```

```
## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## character vector and factor, coercing into character vector
```

```
fatalres_12 <- inner_join(fatalres_12, address_12)
```

```
## Joining, by = c("STATE", "FDID", "INC_DATE", "INC_NO")
```

```
fatalres_13 <- inner_join(fatalres_13, address_13)
```

```
## Joining, by = c("STATE", "FDID", "INC_DATE", "INC_NO")
```

```
## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## character vector and factor, coercing into character vector
```

```
## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## character vector and factor, coercing into character vector
```

```
## Warning in inner_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## character vector and factor, coercing into character vector
```

```
fatalres_14 <- inner_join(fatalres_14, address_14)
```

```
## Joining, by = c("STATE", "FDID", "INC_DATE", "INC_NO")
```

```
#decode the table
code_07 <- read.dbf("NFIRS_2007/codelookup.dbf")

#decode "original area"
code_area <- code_07[code_07$FIELDID == "AREA_ORIG",]
colnames(code_area) <- c("FIELDID", "AREA_ORIG", "ORIG_AREA")

#decode "HEAT_SOURC"
code_heat <- code_07[code_07$FIELDID == "HEAT_SOURC",]
colnames(code_heat) <- c("FIELDID1", "HEAT_SOURC", "HEAT_SOURCE")
#decoe "FIRST_IGN"
code_first <- code_07[code_07$FIELDID == "FIRST_IGN",]
colnames(code_first) <- c("FIELDID2", "FIRST_IGN", "FIRST_IGNITION")
#decode "CAUSE_IGN"
code_cause <- code_07[code_07$FIELDID == "CAUSE_IGN",]
colnames(code_cause) <- c("FIELDID3", "CAUSE_IGN", "CAUSE_IGNITION")


fatalres_07 <- left_join(fatalres_07, code_area) %>% left_join(., code_heat) %>%
  left_join(.,code_first) %>% left_join(.,code_cause)
```

```
## Joining, by = "AREA_ORIG"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
## Joining, by = "HEAT_SOURC"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
## Joining, by = "FIRST_IGN"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
## Joining, by = "CAUSE_IGN"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
```

```
## factors with different levels, coercing to character vector
```

```
fatalres_08 <- left_join(fatalres_08, code_area) %>% left_join(., code_heat) %>%
  left_join(., code_first) %>% left_join(.,code_cause)
```

```
## Joining, by = "AREA_ORIG"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
## Joining, by = "HEAT_SOURC"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
## Joining, by = "FIRST_IGN"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
## Joining, by = "CAUSE_IGN"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
fatalres_09 <- left_join(fatalres_09, code_area)%>% left_join(., code_heat) %>%
  left_join(.,code_first) %>% left_join(.,code_cause)
```

```
## Joining, by = "AREA_ORIG"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
## Joining, by = "HEAT_SOURC"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
## Joining, by = "FIRST_IGN"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
## Joining, by = "CAUSE_IGN"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
fatalres_10 <- left_join(fatalres_10, code_area)%>% left_join(., code_heat) %>%
  left_join(.,code_first) %>% left_join(.,code_cause)
```

```
## Joining, by = "AREA_ORIG"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
## Joining, by = "HEAT_SOURC"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
## Joining, by = "FIRST_IGN"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
## Joining, by = "CAUSE_IGN"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
fatalres_11 <- left_join(fatalres_11, code_area)%>% left_join(., code_heat) %>%
  left_join(.,code_first) %>% left_join(.,code_cause)
```

```
## Joining, by = "AREA_ORIG"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
## Joining, by = "HEAT_SOURC"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
## Joining, by = "FIRST_IGN"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
## Joining, by = "CAUSE_IGN"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factors with different levels, coercing to character vector
```

```
fatalres_12 <- left_join(fatalres_12, code_area)%>% left_join(., code_heat) %>%
  left_join(.,code_first) %>% left_join(.,code_cause)
```

```
## Joining, by = "AREA_ORIG"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factor and character vector, coercing into character vector
```

```
## Joining, by = "HEAT_SOURC"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factor and character vector, coercing into character vector
```

```
## Joining, by = "FIRST_IGN"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factor and character vector, coercing into character vector
```

```
## Joining, by = "CAUSE_IGN"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factor and character vector, coercing into character vector
```

```
fatalres_13 <- left_join(fatalres_13, code_area)%>% left_join(., code_heat) %>%
  left_join(.,code_first) %>% left_join(.,code_cause)
```

```
## Joining, by = "AREA_ORIG"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factor and character vector, coercing into character vector
```

```
## Joining, by = "HEAT_SOURC"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factor and character vector, coercing into character vector
```

```
## Joining, by = "FIRST_IGN"
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factor and character vector, coercing into character vector

## Joining, by = "CAUSE_IGN"

## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factor and character vector, coercing into character vector
```

```
fatalres_14 <- left_join(fatalres_14, code_area)%>% left_join(., code_heat) %>%
  left_join(.,code_first) %>% left_join(.,code_cause)
```

```
## Joining, by = "AREA_ORIG"

## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factor and character vector, coercing into character vector

## Joining, by = "HEAT_SOURC"

## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factor and character vector, coercing into character vector

## Joining, by = "FIRST_IGN"

## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factor and character vector, coercing into character vector

## Joining, by = "CAUSE_IGN"

## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factor and character vector, coercing into character vector
```

```
#remove duplicated columns
colnames(fatalres_07)
```

```
##  [1] "STATE"          "FDID"           "INC_DATE"       "INC_NO"
##  [5] "AREA_ORIG"      "HEAT_SOURC"     "FIRST_IGN"      "CAUSE_IGN"
##  [9] "EXP_NO"         "LOC_TYPE"       "NUM_MILE"       "STREET_PRE"
## [13] "STREETNAME"     "STREETTYPE"     "STREETSUF"      "APT_NO"
## [17] "CITY"           "STATE_ID"       "ZIP5"           "ZIP4"
## [21] "X_STREET"       "FIELDID"        "ORIG_AREA"      "FIELDID1"
## [25] "HEAT_SOURCE"    "FIELDID2"       "FIRST_IGNITION" "FIELDID3"
## [29] "CAUSE_IGNITION"
```

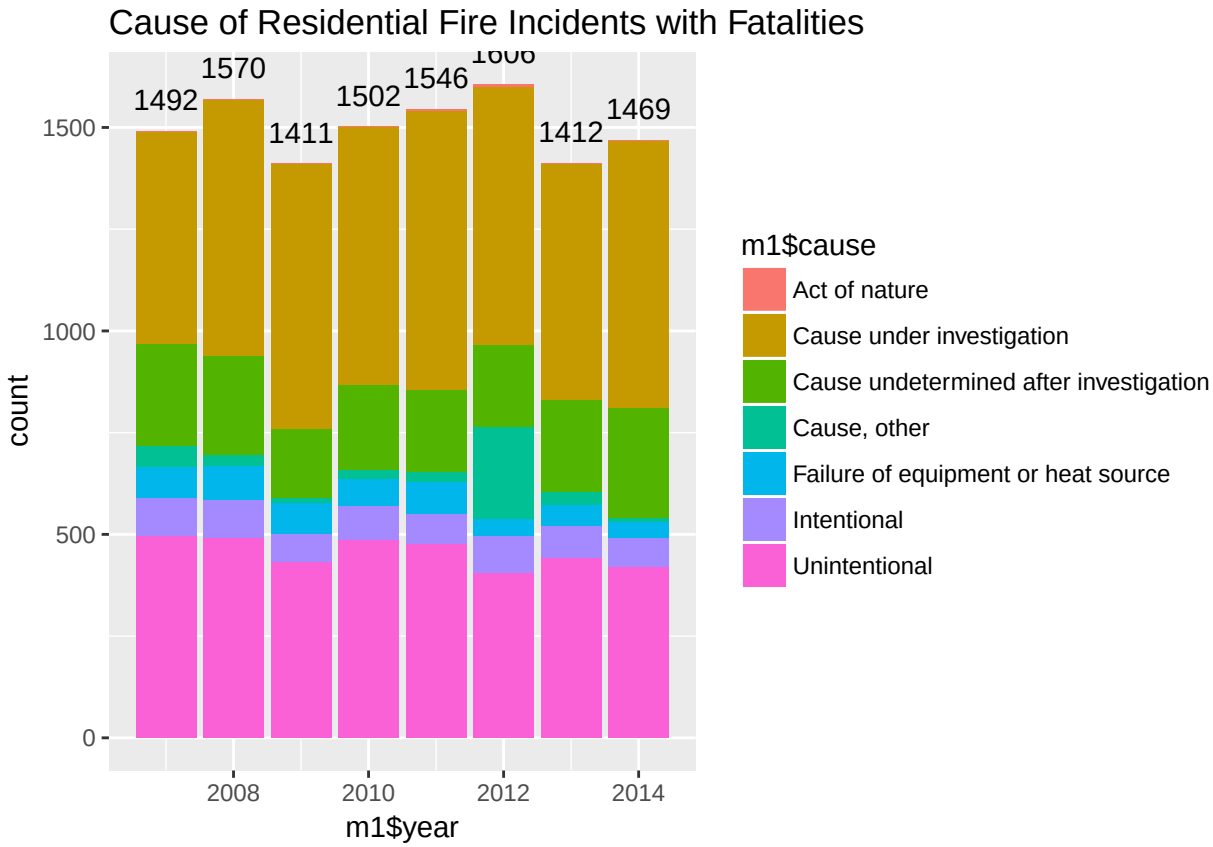
```
fatalres_07 <- fatalres_07[,-c(5:9, 18, 20, 22, 24, 26, 28)]
fatalres_08 <- fatalres_08[,-c(5:9, 18, 20, 22, 24, 26, 28)]
fatalres_09 <- fatalres_09[,-c(5:9, 18, 20, 22, 24, 26, 28)]
fatalres_10 <- fatalres_10[,-c(5:9, 18, 20, 22, 24, 26, 28)]
fatalres_11 <- fatalres_11[,-c(5:9, 18, 20, 22, 24, 26, 28)]
fatalres_12 <- fatalres_12[,-c(5:9, 18, 20, 22, 24, 26, 28)]
fatalres_13 <- fatalres_13[,-c(5:9, 18, 20, 22, 24, 26, 28)]
fatalres_14 <- fatalres_14[,-c(5:9, 18, 20, 22, 24, 26, 28)]

#checked the data, in fatalres_08, there is a bad data, delete it
fatalres_08 <- fatalres_08[-c(107),]


#I made an shiny app in order to see those tables more clearly
# http://127.0.0.1:6356
```

```r
#GRAPHS


#horizontal comparision

#Diane said that causes of fatal residential fires in the period 2011 to 2015
#with the reported causes of fatal residential fires the period 2006 to 2010
#I want to figure out why

#make a table combine years and cause of factions

year07 = as.numeric(fatalres_07$INC_DATE)%%10^4
cause_07 <- data.table(year07,fatalres_07$CAUSE_IGNITION)
colnames(cause_07) <- c("year", "cause")

year08 = as.numeric(fatalres_08$INC_DATE)%%10^4
cause_08 <- data.table(year08,fatalres_08$CAUSE_IGNITION)
colnames(cause_08) <- c("year", "cause")

year09 = as.numeric(fatalres_09$INC_DATE)%%10^4
cause_09 <- data.table(year09,fatalres_09$CAUSE_IGNITION)
colnames(cause_09) <- c("year", "cause")

year10 = as.numeric(fatalres_10$INC_DATE)%%10^4
cause_10 <- data.table(year10,fatalres_10$CAUSE_IGNITION)
colnames(cause_10) <- c("year", "cause")

year11 = as.numeric(fatalres_11$INC_DATE)%%10^4
cause_11 <- data.table(year11,fatalres_11$CAUSE_IGNITION)
colnames(cause_11) <- c("year", "cause")

year12 = as.numeric(fatalres_12$INC_DATE)%%10^4
cause_12 <- data.table(year12,fatalres_12$CAUSE_IGNITION)
colnames(cause_12) <- c("year", "cause")

year13 = as.numeric(fatalres_13$INC_DATE)%%10^4
cause_13 <- data.table(year13,fatalres_13$CAUSE_IGNITION)
colnames(cause_13) <- c("year", "cause")

year14 = as.numeric(fatalres_14$INC_DATE)%%10^4
cause_14 <- data.table(year14,fatalres_14$CAUSE_IGNITION)
colnames(cause_14) <- c("year", "cause")

m1 <- full_join(cause_07, cause_08)%>%full_join(cause_10)%>%full_join(cause_09)%>%
  full_join(cause_11)%>%full_join(cause_12)%>%full_join(cause_13)%>%full_join(cause_14)
```

```
## Joining, by = c("year", "cause")
## Joining, by = c("year", "cause")
## Joining, by = c("year", "cause")
## Joining, by = c("year", "cause")
## Joining, by = c("year", "cause")
## Joining, by = c("year", "cause")
```

```
## Joining, by = c("year", "cause")
```

```
#bar graph of
cause <- ggplot(m1, aes(x =m1$year)) +
  geom_bar(aes(fill = m1$cause), stat = "count")+
  geom_text(stat = 'count', aes(label=..count..),vjust=-1)+
  ggtitle("Cause of Residential Fire Incidents with Fatalities")
cause
```




```
#Analyze of barplots:

#From barplot of Cause of Residential Fire Incidents with Fatalities, we can see that the percentage of
#Diane finds out that the cuase "unknown" is increasing, I am going to have a further look of those inc

#table of unknown cause in fire accidents
unknown <- subset(m1, m1$cause == "Cause under investigation")
u1 <- subset(m1, m1$cause == "Cause undetermined after investigation")
unknown <- full_join(unknown, u1)
```

```
## Joining, by = c("year", "cause")
```

```
head(unknown)
```

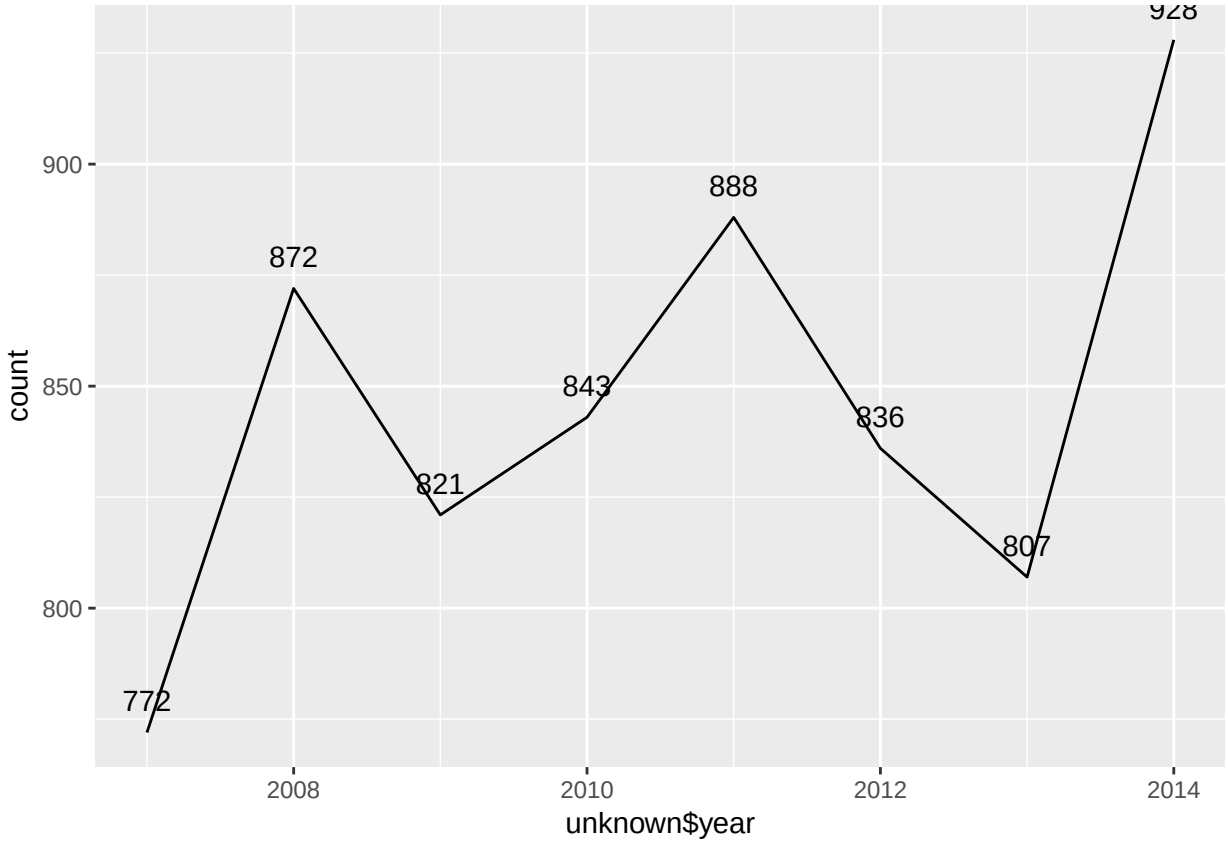
```
##   year                       cause
## 1 2007 Cause under investigation
## 2 2007 Cause under investigation
## 3 2007 Cause under investigation
## 4 2007 Cause under investigation
## 5 2007 Cause under investigation
```

```
## 6 2007 Cause under investigation
```

```
#plot of unknown cause by year
uknownp <- ggplot(unknown, aes(x =unknown$year))+
  geom_text(stat = 'count', aes(label=..count..),vjust=-1)+
  geom_line( stat = "count")
  ggtitle("Unknown Cause of Residential Fire Incidents with Fatalities")
```

```
## $title
## [1] "Unknown Cause of Residential Fire Incidents with Fatalities"
##
## $subtitle
## NULL
##
## attr(,"class")
## [1] "labels"
```

```
uknownp
```




```
#From the plot, we can see that year 2007 and 2013 has the lowest number of incident as cause "unknown"
#It could be explained by that in these two years, the number of fire incidents is also lower than othe


#I decide to conduct a comparison t test to check if there is a difference between two time period


#1.make a table
prop1 <- c(523/1492, 628/1570, 650/1411, 502/1632)
prop2 <- c(688/1546, 634/1606, 581/1412, 658/1469)
```

```r
t1 <- data.frame(cbind(prop1,prop2))
as.factor(t1$prop1)
```

```
## [1] 0.350536193029491 0.4               0.460666194188519 0.307598039215686
## Levels: 0.307598039215686 0.350536193029491 0.4 0.460666194188519
```

```r
as.factor(t1$prop2)
```

```
## [1] 0.445019404915912 0.394769613947696 0.411473087818697 0.447923757658271
## 4 Levels: 0.394769613947696 0.411473087818697 ... 0.447923757658271
```

```r
colnames(t1)<- c("prop of unknown 07-10", "prop of unknown 11-14")
t1
```

```
##   prop of unknown 07-10 prop of unknown 11-14
## 1             0.3505362             0.4450194
## 2             0.4000000             0.3947696
## 3             0.4606662             0.4114731
## 4             0.3075980             0.4479238
```

```r
#comparison t test
t.test(t1$`prop of unknown 07-10`,t1$`prop of unknown 11-14`)
```

```
##
##  Welch Two Sample t-test
##
## data:  t1$`prop of unknown 07-10` and t1$`prop of unknown 11-14`
## t = -1.2738, df = 3.9105, p-value = 0.2732
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.14428248  0.05408976
## sample estimates:
## mean of x mean of y
## 0.3797001 0.4247965
```

```r
#RESULT

#Welch Two Sample t-test

#data:  t1$prop1 and t1$prop2
#t = -1.2738, df = 3.9105, p-value = 0.2732
#alternative hypothesis: true difference in means is not equal to 0
#95 percent confidence interval:
#  -0.14428248  0.05408976
#sample estimates:
#  mean of x   mean of y
# 0.3797001    0.4247965


#ANALYSIS
#Since the p-value = 0.2732, which is bigger than 0.05, I conclude that there si no difference between
#The unknown of cause of fatal residential fires which is quite stable in ten years. Probably Diane  di
```

```
#MAPS(script in another file)



#Analyze from map:

#From the map, we can see that fire incidence with fatalities often happens in state CA, New York, Flor
#It is understandable, because in those places, there are more people live there. Therefore, the occura

#Shiny APP http://127.0.0.1:6356
```