NFIRS Arson Casualties

Emma Brown

This project investigates casualties due to arson in the United States from 2010 to 2015.

The data comes from the National Fire Incident Reporting System (NFIRS). The data was cleaned and organized by time data, such as month and year, demographic data, such as gender and race, and location data, including city, state, and zip code.

Of the deliverable files, the relevant files to this project were arson, civiliancasualty, and incidentaddress. Code lookup was also included to ensure coding consistency across dbf files.

Libraries were loaded

```
require(dplyr)
```

```
## Loading required package: 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
```

```
require(tidyr)
```

```
## Loading required package: tidyr
```

```
require(foreign)
```

```
## Loading required package: foreign
```

```
require(data.table)
```

```
## Loading required package: 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
```

```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
require(maps)
```

```
## Loading required package: maps
```

```
require(mapproj)
```

```
## Loading required package: mapproj
```

```
require(devtools)
```

```
## Loading required package: devtools
```

```
require(choroplethr)
```

```
## Loading required package: choroplethr
```

```
require(stringr)
```

```
## Loading required package: stringr
```

```
require(zipcode)
```

```
## Loading required package: zipcode
```

```
require(ggmap)
```

```
## Loading required package: ggmap
```

```
require(knitr)
```

```
## Loading required package: knitr
```

```
require(reshape2)
```

```
## Loading required package: reshape2
```

```
##
## Attaching package: 'reshape2'
```

```
## The following objects are masked from 'package:data.table':
##
##     dcast, melt
```

```
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```
data(zipcode)
data(state)
```

I first looked at 2010.

```
# Load the relevant files
arson <- read.dbf("arson2010.dbf")
arson$YEAR <- 2010
code_lookup <- read.dbf("codelookup2010.dbf")
incidentaddress <- read.dbf("incidentaddress2010.dbf")
civiliancasualty <- read.dbf("civiliancasualty2010.dbf")
civiliancasualty$YEAR <- 2010
```

Now let's start cleaning. Some relevant attributes were selected. Including ID attributes to join databases.

```
civiliancasualty <- civiliancasualty %>%
  select(YEAR, STATE, FDID, INC_DATE, INC_NO, GENDER, AGE, RACE, SEV)
```

```
arson <- arson %>%
  select(YEAR, STATE, FDID, INC_DATE, INC_NO)
```

Changed the levels of civilian casualties according to their code lookup classification. Got rid of any unknowns because we're only interested in injuries and casualties. Then changed the race and gender levels according to code lookup for better looking data frame.

```
levels(civiliancasualty$SEV) <- c("Mild",
                                  "Moderate",
                                  "Severe",
                                  "Life Threatening",
                                  "Death",
                                  "Unknown")
# Remove Unknowns
civiliancasualty <- civiliancasualty %>%
  filter(SEV != "Unknown")
# Race Levels
levels(civiliancasualty$RACE) <- c("NA",
                                   "White",
                                   "Black",
                                   "American Indian",
                                   "Asian",
                                   "Native Hawaiian",
                                   "NA")
# Gender Levels
levels(civiliancasualty$GENDER) <- c("Male",
                                     "Female")
```

Now let's start putting some data together... I joined the arson and civilian casualty data frames to find the civilian casualty from arson. Then joined that data frame with incident address to find the location of arson incidents

```
# joins the casualty dbf with arson dbf to find civilian casualties from arson
arsoncasualties <- semi_join(civiliancasualty, arson,
                             by=c('INC_NO'='INC_NO', 'STATE'='STATE',
                                  'INC_DATE'='INC_DATE'))

# joins to find incident addresses
arsoncasualties <- left_join(arsoncasualties, incidentaddress)
```

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

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

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

```
arsoncasualties <- arsoncasualties %>%
  select(YEAR, STATE, FDID, INC_DATE, INC_NO, GENDER,
         AGE, RACE, SEV, STREETNAME, STREETTYPE,
         CITY, ZIP=ZIP5)
```

To look at some of this data, I made a data frame of frequencies of arson incidents with casualties by state.

```
# Frequency by state
arsonbystate <- arsoncasualties %>%
```

```
  group_by(STATE) %>%
  summarise(CASUALTIES=n())
```

Then I reformatted the dates to find the month of each incident. Then I saved the data frame to its respective year to combine later.

```
arsoncasualties$INC_DATE <- str_pad(arsoncasualties$INC_DATE, 8, pad="0")
arsoncasualties <- transform(arsoncasualties, INC_DATE = as.Date(as.character(INC_DATE), "%m%d%Y"))
arsoncasualties$MONTH <- format(as.Date(arsoncasualties$INC_DATE, "%m"))
arsoncasualties$MONTH <- substr(arsoncasualties$INC_DATE, 6, 7)
arsoncasualties$MONTH <- as.integer(arsoncasualties$MONTH)
arsoncasualties$MONTH <- month.abb[arsoncasualties$MONTH]
arsoncasualties_2010 <- arsoncasualties
```

Now for 2011:

```
arson <- read.dbf("arson2011.dbf")
arson$YEAR <- 2011
code_lookup <- read.dbf("codelookup2011.dbf")
incidentaddress <- read.dbf("incidentaddress2011.dbf")

# Civilian Casualties
civiliancasualty <- read.dbf("civilian2011.dbf")
# Select Column Names
civiliancasualty$YEAR <- 2011
civiliancasualty <- civiliancasualty %>%
  select(YEAR, STATE, FDID, INC_DATE, INC_NO, GENDER, AGE, RACE, SEV)
# Select Column Names
arson <- arson %>%
  select(YEAR, STATE, FDID, INC_DATE, INC_NO)

# Change Severity levels
levels(civiliancasualty$SEV) <- c("Mild",
                                  "Moderate",
                                  "Severe",
                                  "Life Threatening",
                                  "Death",
                                  "Unknown")
# Remove Unknowns
civiliancasualty <- civiliancasualty %>%
  filter(SEV != "Unknown")
# Race Levels
levels(civiliancasualty$RACE) <- c("NA",
                                   "White",
                                   "Black",
                                   "American Indian",
                                   "Asian",
                                   "Native Hawaiian",
                                   "NA")
# Gender Levels
levels(civiliancasualty$GENDER) <- c("Male",
                                     "Female")
# joins the casualty dbf with arson dbf to find civilian casualties from arson
arsoncasualties <- semi_join(civiliancasualty, arson,
                             by=c('INC_NO'='INC_NO', 'STATE'='STATE',
```

```
                                    'INC_DATE'='INC_DATE'))

# joins to find incident addresses
arsoncasualties <- left_join(arsoncasualties, incidentaddress)
```

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

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

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

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

```
arsoncasualties <- arsoncasualties %>%
  select(YEAR, STATE, FDID, INC_DATE, INC_NO, GENDER,
         AGE, RACE, SEV, STREETNAME, STREETTYPE,
         CITY, ZIP=ZIP5)

# Frequency by state
arsonbystate <- arsoncasualties %>%
  group_by(STATE) %>%
  summarise(count=n())

arsoncasualties$INC_DATE <- str_pad(arsoncasualties$INC_DATE, 8, pad="0")
arsoncasualties <- transform(arsoncasualties, INC_DATE = as.Date(as.character(INC_DATE), "%m%d%Y"))
arsoncasualties$MONTH <- format(as.Date(arsoncasualties$INC_DATE, "%m"))
arsoncasualties$MONTH <- substr(arsoncasualties$INC_DATE, 6, 7)
arsoncasualties$MONTH <- as.integer(arsoncasualties$MONTH)
arsoncasualties$MONTH <- month.abb[arsoncasualties$MONTH]
arsoncasualties_2011 <- arsoncasualties
```

On to 2012: With 2012 onward, the dataframes were in text files, which means we have to open them differently while perserving the headings and data, but the process is still generally the same.

```
arson <- read.table("arson2012.txt",
                    header=TRUE,
                    sep="^")
arson$YEAR <- 2012
code_lookup <- read.table("codelookup2012.txt",
                          header=TRUE,
                          sep="^")
incidentaddress <- read.table("incidentaddress2012.txt",
                              header=TRUE,
                              sep="^")

# Civilian Casualties
civiliancasualty <- read.table("civiliancasualty2012.txt",
                               header=TRUE,
                               sep="^")
# Select Column Names
civiliancasualty$YEAR <- 2012
civiliancasualty <- civiliancasualty %>%
```

```
  select(YEAR, STATE, FDID, INC_DATE, INC_NO, GENDER, AGE, RACE, SEV)
# Select Column Names
arson <- arson %>%
  select(YEAR, STATE, FDID, INC_DATE, INC_NO)

# Change Severity levels
levels(civiliancasualty$SEV) <- c("Mild",
                                  "Moderate",
                                  "Severe",
                                  "Life Threatening",
                                  "Death",
                                  "Unknown")
# Remove Unknowns
civiliancasualty <- civiliancasualty %>%
  filter(SEV != "Unknown")
# Race Levels
levels(civiliancasualty$RACE) <- c("NA",
                                   "NA",
                                   "White",
                                   "Black",
                                   "American Indian",
                                   "Asian",
                                   "Native Hawaiian",
                                   "NA")
# Gender Levels
civiliancasualty$GENDER <- as.factor(civiliancasualty$GENDER)
levels(civiliancasualty$GENDER) <- c("Male",
                                     "Female")
# joins the casualty dbf with arson dbf to find civilian casualties from arson
arsoncasualties <- semi_join(civiliancasualty, arson,
                             by=c('INC_NO'='INC_NO', 'STATE'='STATE',
                                  'INC_DATE'='INC_DATE'))

# joins to find incident addresses
arsoncasualties <- left_join(arsoncasualties, incidentaddress)
```

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

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

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

```
arsoncasualties <- arsoncasualties %>%
  select(YEAR, STATE, FDID, INC_DATE, INC_NO, GENDER,
         AGE, RACE, SEV, STREETNAME, STREETTYPE,
         CITY, ZIP=ZIP5)

# Frequency by state
arsonbystate <- arsoncasualties %>%
  group_by(STATE) %>%
  summarise(count=n())

arsoncasualties$INC_DATE <- str_pad(arsoncasualties$INC_DATE, 8, pad="0")
```

```
arsoncasualties <- transform(arsoncasualties, INC_DATE = as.Date(as.character(INC_DATE), "%m%d%Y"))
arsoncasualties$MONTH <- format(as.Date(arsoncasualties$INC_DATE, "%m"))
arsoncasualties$MONTH <- substr(arsoncasualties$INC_DATE, 6, 7)
arsoncasualties$MONTH <- as.integer(arsoncasualties$MONTH)
arsoncasualties$MONTH <- month.abb[arsoncasualties$MONTH]
arsoncasualties_2012 <- arsoncasualties
```

And 2013:

```
arson <- read.table("arson2013.txt",
                    header=TRUE,
                    sep="^")
arson$YEAR <- 2013
code_lookup <- read.table("codelookup2013.txt",
                          header=TRUE,
                          sep="^")
incidentaddress <- read.table("incidentaddress2013.txt",
                              header=TRUE,
                              sep="^")
```

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

```
# Civilian Casualties
civiliancasualty <- read.table("civiliancasualty2013.txt",
                               header=TRUE,
                               sep="^")
# Select Column Names
civiliancasualty$YEAR <- 2013
civiliancasualty <- civiliancasualty %>%
  select(YEAR, STATE, FDID, INC_DATE, INC_NO, GENDER, AGE, RACE, SEV)
# Select Column Names
arson <- arson %>%
  select(YEAR, STATE, FDID, INC_DATE, INC_NO)

# Change Severity levels
levels(civiliancasualty$SEV) <- c("NA",
                                  "Mild",
                                  "Moderate",
                                  "Severe",
                                  "Life Threatening",
                                  "Death",
                                  "Unknown")
# Remove Unknowns
civiliancasualty <- civiliancasualty %>%
  filter(SEV != "Unknown",
         SEV != "NA")
# Race Levels
levels(civiliancasualty$RACE) <- c("NA",
                                   "NA",
                                   "White",
                                   "Black",
                                   "American Indian",
                                   "Asian",
                                   "Native Hawaiian",
```

```
                                    "NA")
# Gender Levels
civiliancasualty$GENDER <- as.factor(civiliancasualty$GENDER)
levels(civiliancasualty$GENDER) <- c("Male",
                                     "Female")
# joins the casualty dbf with arson dbf to find civilian casualties from arson
arsoncasualties <- semi_join(civiliancasualty, arson,
                             by=c('INC_NO'='INC_NO', 'STATE'='STATE',
                                  'INC_DATE'='INC_DATE'))

# joins to find incident addresses
arsoncasualties <- left_join(arsoncasualties, incidentaddress)
```

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

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

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

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

```
arsoncasualties <- arsoncasualties %>%
  select(YEAR, STATE, FDID, INC_DATE, INC_NO, GENDER,
         AGE, RACE, SEV, STREETNAME, STREETTYPE,
         CITY, ZIP=ZIP5)

# Frequency by state
arsonbystate <- arsoncasualties %>%
  group_by(STATE) %>%
  summarise(count=n())

arsoncasualties$INC_DATE <- str_pad(arsoncasualties$INC_DATE, 8, pad="0")
arsoncasualties <- transform(arsoncasualties, INC_DATE = as.Date(as.character(INC_DATE), "%m%d%Y"))
arsoncasualties$MONTH <- format(as.Date(arsoncasualties$INC_DATE, "%m"))
arsoncasualties$MONTH <- substr(arsoncasualties$INC_DATE, 6, 7)
arsoncasualties$MONTH <- as.integer(arsoncasualties$MONTH)
arsoncasualties$MONTH <- month.abb[arsoncasualties$MONTH]
arsoncasualties_2013 <- arsoncasualties
```

2014:

```
arson <- read.table("arson2014.txt",
                    header=TRUE,
                    sep="^")
arson$YEAR <- 2014
code_lookup <- read.table("codelookup2014.txt",
                          header=TRUE,
                          sep="^")
incidentaddress <- read.table("incidentaddress2014.txt",
                              header=TRUE,
                              sep="^")

# Civilian Casualties
```

```
civiliancasualty <- read.table("civiliancasualty2014.txt",
                               header=TRUE,
                               sep="^")
# Select Column Names
civiliancasualty$YEAR <- 2014
civiliancasualty <- civiliancasualty %>%
  select(YEAR, STATE, FDID, INC_DATE, INC_NO, GENDER, AGE, RACE, SEV)
# Select Column Names
arson <- arson %>%
  select(YEAR, STATE, FDID, INC_DATE, INC_NO)

# Change Severity levels
levels(civiliancasualty$SEV) <- c("Mild",
                                  "Moderate",
                                  "Severe",
                                  "Life Threatening",
                                  "Death",
                                  "Unknown")
# Remove Unknowns
civiliancasualty <- civiliancasualty %>%
  filter(SEV != "Unknown")
# Race Levels
levels(civiliancasualty$RACE) <- c("NA",
                                   "NA",
                                   "White",
                                   "Black",
                                   "American Indian",
                                   "Asian",
                                   "Native Hawaiian",
                                   "NA")
# Gender Levels
civiliancasualty$GENDER <- as.factor(civiliancasualty$GENDER)
levels(civiliancasualty$GENDER) <- c("Male",
                                     "Female")
# joins the casualty dbf with arson dbf to find civilian casualties from arson
arsoncasualties <- semi_join(civiliancasualty, arson,
                             by=c('INC_NO'='INC_NO', 'STATE'='STATE',
                                  'INC_DATE'='INC_DATE'))

# joins to find incident addresses
arsoncasualties <- left_join(arsoncasualties, incidentaddress)
```

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

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

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

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

```
arsoncasualties <- arsoncasualties %>%
  select(YEAR, STATE, FDID, INC_DATE, INC_NO, GENDER,
         AGE, RACE, SEV, STREETNAME, STREETTYPE,
         CITY, ZIP=ZIP5)

# Frequency by state
arsonbystate <- arsoncasualties %>%
  group_by(STATE) %>%
  summarise(count=n())

arsoncasualties$INC_DATE <- str_pad(arsoncasualties$INC_DATE, 8, pad="0")
arsoncasualties <- transform(arsoncasualties, INC_DATE = as.Date(as.character(INC_DATE), "%m%d%Y"))
arsoncasualties$MONTH <- format(as.Date(arsoncasualties$INC_DATE, "%m"))
arsoncasualties$MONTH <- substr(arsoncasualties$INC_DATE, 6, 7)
arsoncasualties$MONTH <- as.integer(arsoncasualties$MONTH)
arsoncasualties$MONTH <- month.abb[arsoncasualties$MONTH]
arsoncasualties_2014 <- arsoncasualties
```

And 2015:

```
arson <- read.table("arson2015.txt",
                    header=TRUE,
                    sep="^")
arson$YEAR <- 2015
code_lookup <- read.table("codelookup2015.txt",
                          header=TRUE,
                          sep="^")
incidentaddress <- read.table("incidentaddress2015.txt",
                              header=TRUE,
                              sep="^")

# Civilian Casualties
civiliancasualty <- read.table("civiliancasualty2015.txt",
                               header=TRUE,
                               sep="^")
# Select Column Names
civiliancasualty$YEAR <- 2015
civiliancasualty <- civiliancasualty %>%
  select(YEAR, STATE, FDID, INC_DATE, INC_NO, GENDER, AGE, RACE, SEV)
# Select Column Names
arson <- arson %>%
  select(YEAR, STATE, FDID, INC_DATE, INC_NO)

# Change Severity levels
levels(civiliancasualty$SEV) <- c("Mild",
                                  "Moderate",
                                  "Severe",
                                  "Life Threatening",
                                  "Death",
                                  "Unknown")
# Remove Unknowns
civiliancasualty <- civiliancasualty %>%
  filter(SEV != "Unknown")
# Race Levels
```

```
levels(civiliancasualty$RACE) <- c("NA",
                                   "NA",
                                   "White",
                                   "Black",
                                   "American Indian",
                                   "Asian",
                                   "Native Hawaiian",
                                   "NA")
# Gender Levels
civiliancasualty$GENDER <- as.factor(civiliancasualty$GENDER)
levels(civiliancasualty$GENDER) <- c("Male",
                                     "Female")
# joins the casualty dbf with arson dbf to find civilian casualties from arson
arsoncasualties <- semi_join(civiliancasualty, arson,
                             by=c('INC_NO'='INC_NO', 'STATE'='STATE',
                                  'INC_DATE'='INC_DATE'))

# joins to find incident addresses
arsoncasualties <- left_join(arsoncasualties, incidentaddress)
```

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

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

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

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

```
arsoncasualties <- arsoncasualties %>%
  select(YEAR, STATE, FDID, INC_DATE, INC_NO, GENDER,
         AGE, RACE, SEV, STREETNAME, STREETTYPE,
         CITY, ZIP=ZIP5)

# Frequency by state
arsonbystate <- arsoncasualties %>%
  group_by(STATE) %>%
  summarise(count=n())

arsoncasualties$INC_DATE <- str_pad(arsoncasualties$INC_DATE, 8, pad="0")
arsoncasualties <- transform(arsoncasualties, INC_DATE = as.Date(as.character(INC_DATE), "%m%d%Y"))
arsoncasualties$MONTH <- format(as.Date(arsoncasualties$INC_DATE, "%m"))
arsoncasualties$MONTH <- substr(arsoncasualties$INC_DATE, 6, 7)
arsoncasualties$MONTH <- as.integer(arsoncasualties$MONTH)
arsoncasualties$MONTH <- month.abb[arsoncasualties$MONTH]
arsoncasualties_2015 <- arsoncasualties
```

Finally, we combine all data frames into one.

```
arson_2010_2015 <- do.call(rbind,
                                   list(arsoncasualties_2010,
                                        arsoncasualties_2011,
                                        arsoncasualties_2012,
```

```
                                          arsoncasualties_2013,
                                          arsoncasualties_2014,
                                          arsoncasualties_2015))
```

Make sure they're unique. Then join the zipcode data to find the longitude and latitude of incident addresses. Select the final columns.

```
arson_2010_2015 <- unique(arson_2010_2015)
arson_2010_2015 <- left_join(arson_2010_2015, zipcode, by=c("ZIP"="zip"))
```

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

```
arson_2010_2015 <- arson_2010_2015 %>%
  select(YEAR, MONTH, FDID, INC_DATE, GENDER, AGE, RACE,
         SEV, CITY, STATE=state, ZIP, LAT=latitude,LON=longitude)
```

Removes the data frames from individual years from the workspace to avoid clutter.

```
rm(arsoncasualties_2010,
   arsoncasualties_2011,
   arsoncasualties_2012,
   arsoncasualties_2013,
   arsoncasualties_2014,
   arsoncasualties_2015)
```

Mapping all the casualties in the dataset

```
map<-get_map(location='united states', zoom=4, maptype = "terrain",
             source='google',color='color')
```

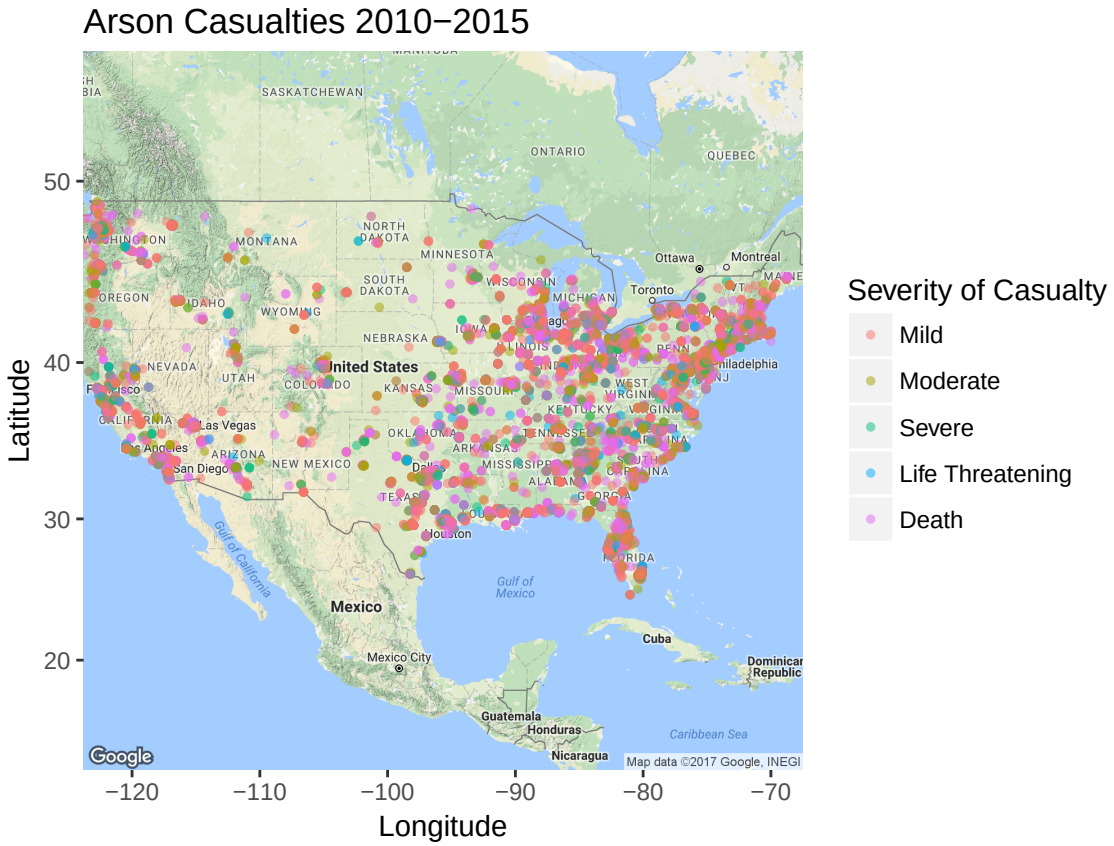
```
## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=united+states&zoom=4&size=640x64
```

```
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=united%20states&sens
```

```
map <- ggmap(map)

map + geom_point(aes(x=LON, y=LAT, color=SEV), data=arson_2010_2015,
                 alpha=.5, size=1) +
  labs(x="Longitude", y="Latitude") +
  ggtitle("Arson Casualties 2010-2015") +
  scale_color_discrete(name="Severity of Casualty")
```

```
## Warning: Removed 329 rows containing missing values (geom_point).
```



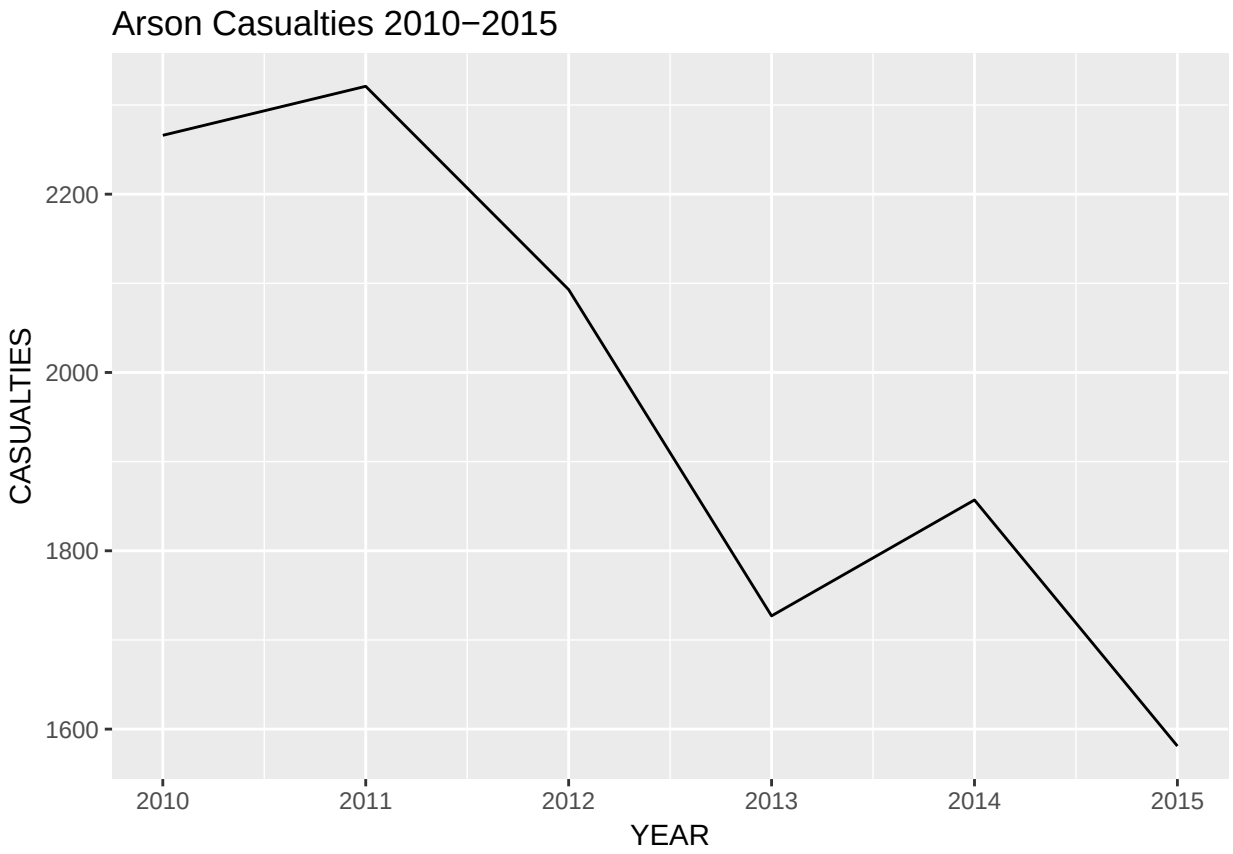
Let's group these to find the frequency of casualties for each year and make a table to display it

```
arsonbyyear <- arson_2010_2015 %>%
  group_by(YEAR) %>%
  summarise(CASUALTIES=n()) %>%
  mutate(PERCENT=CASUALTIES/sum(CASUALTIES))
kable(arsonbyyear)
```

YEAR	CASUALTIES	PERCENT
2010	2266	0.1913043
2011	2321	0.1959477
2012	2093	0.1766990
2013	1727	0.1457999
2014	1857	0.1567750
2015	1581	0.1334740

Now let's graph it

```
ggplot(arsonbyyear, aes(YEAR, CASUALTIES)) +
  geom_line() +
  ggtitle("Arson Casualties 2010-2015")
```

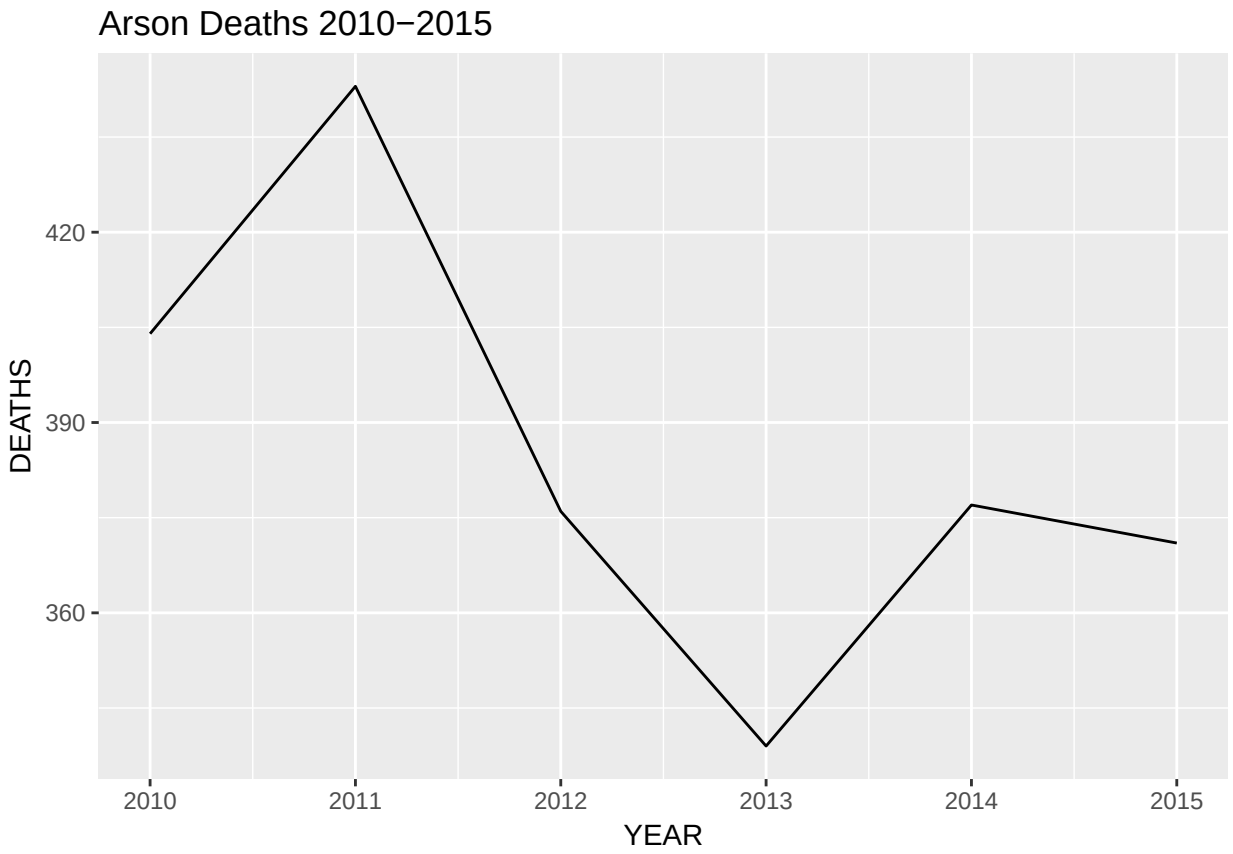


Casualties have certainly gone down since 2010, but what about just casualties resulting in death?

```
arsondeathsbyyear <- arson_2010_2015 %>%
  filter(SEV=='Death') %>%
  group_by(YEAR) %>%
  summarise(DEATHS=n())
kable(arsondeathsbyyear)
```

YEAR	DEATHS
2010	404
2011	443
2012	376
2013	339
2014	377
2015	371

```
ggplot(arsondeathsbyyear, aes(YEAR, DEATHS)) +
  geom_line() +
  ggtitle("Arson Deaths 2010-2015")
```




```
write.csv(arsondeathsbyyear, "arsondeathsbyyear.csv",
          quote=FALSE, row.names=FALSE)
```

There was a spike in 2011, and a significant decrease in 2013, but it has increased again.

```
arsondeathsbystate <- arson_2010_2015 %>%
  filter(SEV=='Death') %>%
  group_by(YEAR,
           STATE) %>%
  summarise(DEATHS=n())

write.csv(arsondeathsbystate, "arsondeathsbystate.csv",
          quote=FALSE, row.names=FALSE)
```