

Firefighter Fatalities, 2000 - 2015

Dan Sisco

10/25/2017

```
library(tidyverse)
library(lubridate)
library(forcats)
library(RColorBrewer)
```

Executive Summary

When I first sat down with this data set, there were three questions I set out to answer before digging into it:

1. What are the top 10 reasons for firefighter fatalities?
2. Are there months where deaths are higher?
3. Are deaths going up or down, overall?

As it turned out, only the first question yielded anything particularly compelling. But as I started going through the data, other stories emerged. I found that Age and Cause of Death are highly correlated, and there are a lot of interesting stories there. One such story is the fact that younger firefighters are more prone to death by “Trauma,” while older firefighters (50+) tend to die of heart attacks much more frequently—though the frequency of firefighter heart attacks on the job is roughly equivalent to the general population. So, as it turned out, “Trauma” was where the really story lies.

I also found that death by “Suffocation” has been on the decline in recent years. I would have liked to dive into that more to see if I could find any information about improvements in firefighter technology or equipment, or anything else that might have caused the decline.

The final graphic shows the Age/Cause histogram; it shows how natural disasters can have a large impact on the number of firefighter fatalities; it shows that “Trauma” is the largest cause of death, by far, after heart attacks; and it shows that Suffocation fatalities are declining.

Data Background

I downloaded the data set, “Firefighter Fatalities in the United States,” from kaggle.com. It details the number of firefighter fatalities by date, firefighter rank, age, incident type, cause of death, and several other variables, across the entire United States from the year 2000 through 2015.

Data Cleaning

This data was relatively well-organized when I got it, so I didn’t have to do a ton to reorganize and clean it. I did, however, parse the dates using lubridate and then mutated those columns into new Year and Month columns so that I could sort and organize the data by those variables.

There were also some categories of firefighter incidents that I filtered out in some cases, such as “NA” and “Other,” because they didn’t add anything to the overall story. In three out of the four graphics I created, I also filtered out fatalities from September 11 so that they didn’t skew the data in any way. Only where it was contextually relevant did I leave the September 11 data in.

Right up front, I needed to rename a number of categories to single-word variables, as shown below, for ease of working with the data. I also created a color palette and put it in this Chunk, which is probably not best practice, but I didn't know where else to put it.

```
Firefighters <- read_csv("data/database.csv") %>%
  rename(Death_Date = `Date of Death`, Incident_Date = `Date of Incident`, Cause = `Cause Of Death`, Na

newPalette <- c("#67001f", "#d53e4f", "#f46d43", "#fdae61", "#fee08b", "#ffffbf", "#e6f598", "#abdda4",
```

Individual Figures

Figure 1: Histogram

The most compelling chart I created was the Age/Cause of Death histogram. As it turned out, it's just a column chart that looks like a histogram, but I'm still calling it a histogram for the purposes of this assignment.

To create this chart, I built a new data frame that included the incident Year, filtered out September 11 and "NA" variables, and then used the Summarise function to count the number of each fatality cause. I plotted Age on the x-axis and Cause of Death on the y-axis, which yielded some really interesting results.

From a design perspective, I actually created a new Palette with colors that made the chart look like a forest on fire, which I was really proud of. You can see this in the final document, but I have since had to alter the Palette in order to facilitate knitting to PDF.

I used the Minimal theme and got rid of the grid lines, because I find them distracting. I put the legend below the graph. In the final product, I made sure to align the title to the left-most side of the chart itself, and I used colors consistently across the entire graphic (mapped to a single legend).

```
#How Are Age and Cause of Death Related?
```

```
ff_age <- Firefighters %>%
  mutate(Incident_Date = mdy(Incident_Date)) %>%
  mutate(Incident_Month = month(Incident_Date, label = TRUE)) %>%
  mutate(Incident_Year = year(Incident_Date)) %>%
  filter(!Incident_Date == "2001-09-11") %>%
  mutate(Age = as.numeric(Age)) %>%
  filter(!Nature %in% c("Unknown", "Other")) %>%
  filter(!is.na(Nature)) %>%
  filter(!is.na(Age)) %>%
  group_by(Age, Nature) %>%
  summarise(Total = n())

age <- ggplot(ff_age, aes(x = Age)) +
  geom_col(aes(y = Total, fill = Nature)) +
  theme_minimal() +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), legend.position = "bottom") +
  scale_fill_manual(values=newPalette)
```

```
age
```

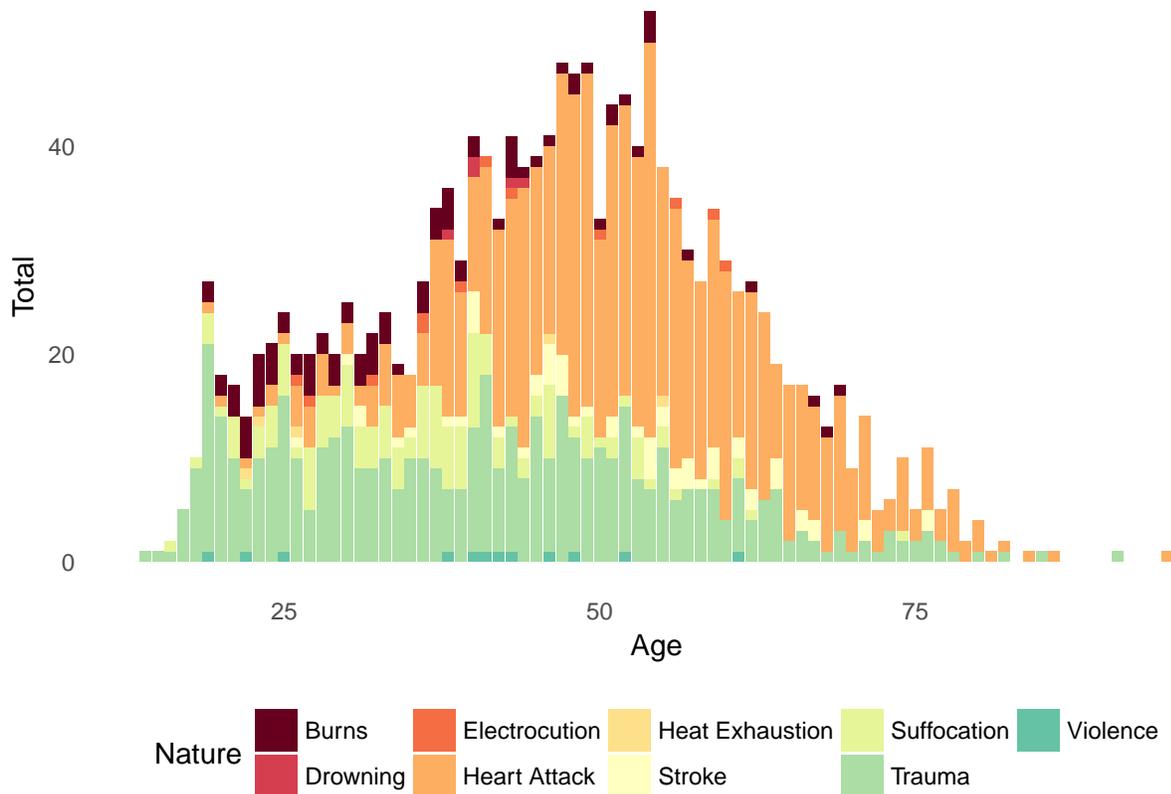


Figure 2: Stacked Column

This graphic was trickier because it isn't as immediately obvious what story it's telling. That's why it didn't make as much sense until I put it in Illustrator and added labels calling out September 11 and the Yarnell Hill Fire in 2013. The point of the graphic is that natural disasters and tragedies can really skew the data—which is why I left the 9/11 data in for this particular chart.

I used the Minimal theme for all of the charts I created, and removed the gridlines, for consistency.

```
#How Have Causes of Fatalities Changed Over Time?

ff_clean <- Firefighters %>%
  mutate(Death_Date = mdy(Death_Date)) %>%
  mutate(Death_Month = month(Death_Date, label = TRUE)) %>%
  mutate(Death_Year = year(Death_Date)) %>%
  mutate(Incident_Date = mdy(Incident_Date)) %>%
  mutate(Incident_Month = month(Incident_Date, label = TRUE)) %>%
  mutate(Incident_Year = year(Incident_Date)) %>%
  filter(!Nature == "Other") %>%
  group_by(Death_Year, Nature) %>%
  count(Nature)

causes_year <- ggplot(ff_clean, aes(x = Death_Year, y = n, fill = Nature)) +
  geom_col() +
  labs(x = "Year of Death", y = "Count", title = "Causes of Firefighter Deaths Since 2000") +
  theme_minimal() +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
  panel.background = element_blank()) +
```

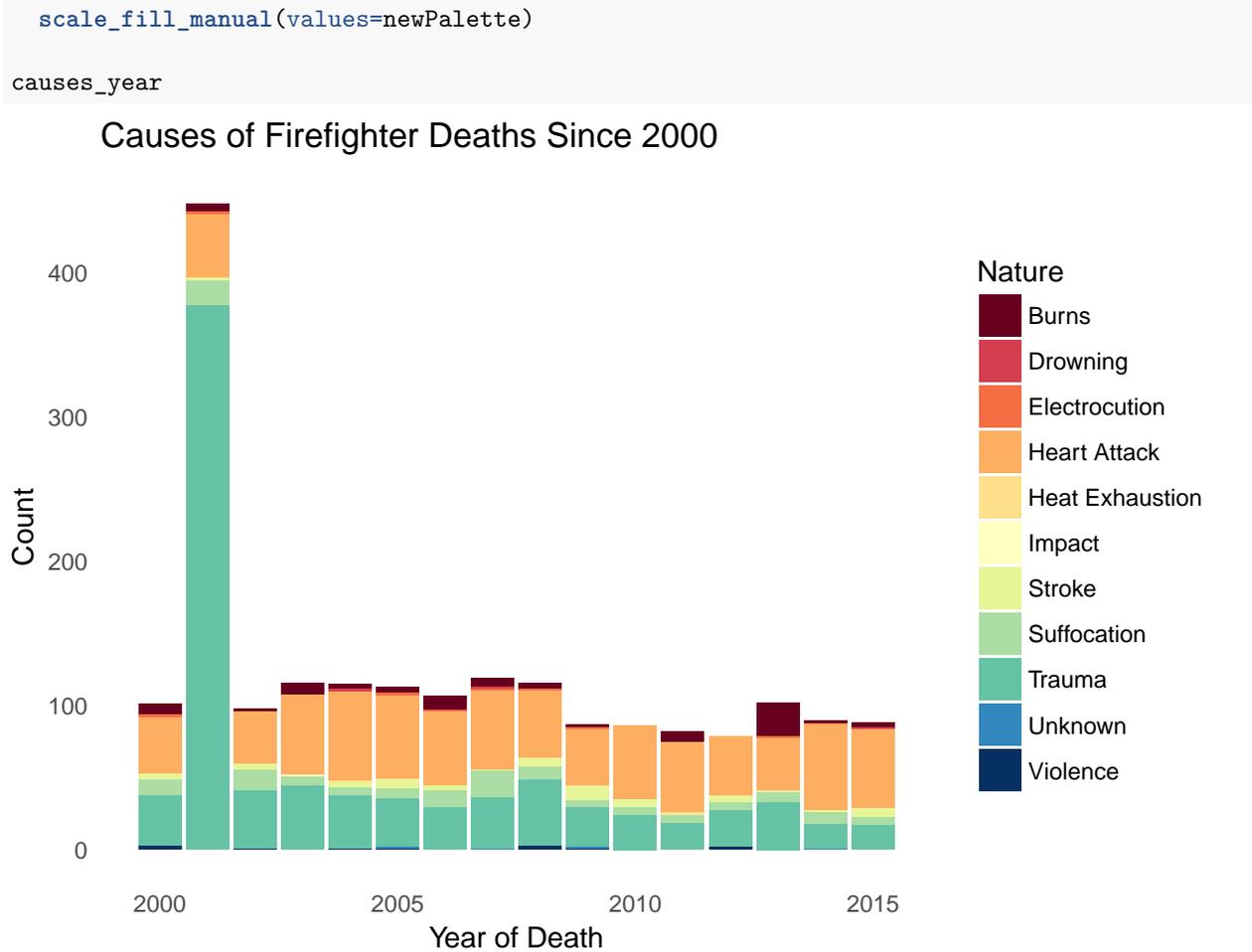


Figure 3: Bar Chart

This is the most basic of the four charts I created. In the final document, I made sure to map the right colors to the right Causes of Death in all of the text, as an element of Repetition (e.g. always highlighting “Trauma” in the same green as is found in the legend).

This chart simply shows the cumulative total of fatalities by cause. Because Heart Attacks are fairly common, I highlighted “Trauma” in the final document, because that is a cause of death that is unique to firefighters.

I like the simplicity of this chart, and the contrast and pop of the green against the rest of the gray bars.

Throughout the final visualization graphic, I made sure to only use dark gray instead of black, because it’s lighter and more interesting to look at than pitch black bars and lines.

I kept getting an error that it “failed to Parse” 1 line. I’m not sure what that line is, I wasn’t able to identify it.

```
#What Are the Top 10 Causes of Fatalities?

ff_causes = Firefighters %>%
  mutate(Incident_Date = mdy(Incident_Date)) %>%
  mutate(Incident_Year = year(Incident_Date)) %>%
  filter(!Incident_Date == "2001-09-11") %>%
  group_by(Nature) %>%
```

```

summarise(Total_Causes = n()) %>%
ungroup() %>%
arrange(desc(Total_Causes)) %>%
slice(1:10) %>%
mutate(Nature = fct_rev(fct_inorder(Nature, ordered = TRUE)))

## Warning: 1 failed to parse.

causes <- ggplot(ff_causes, aes(x = Nature, y = Total_Causes)) +
  geom_col() +
  coord_flip() +
  theme_minimal() +
  labs(y = "Total Number of Incidents", x = "Nature of Incident") +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), legend.position = "bottom")

causes

```

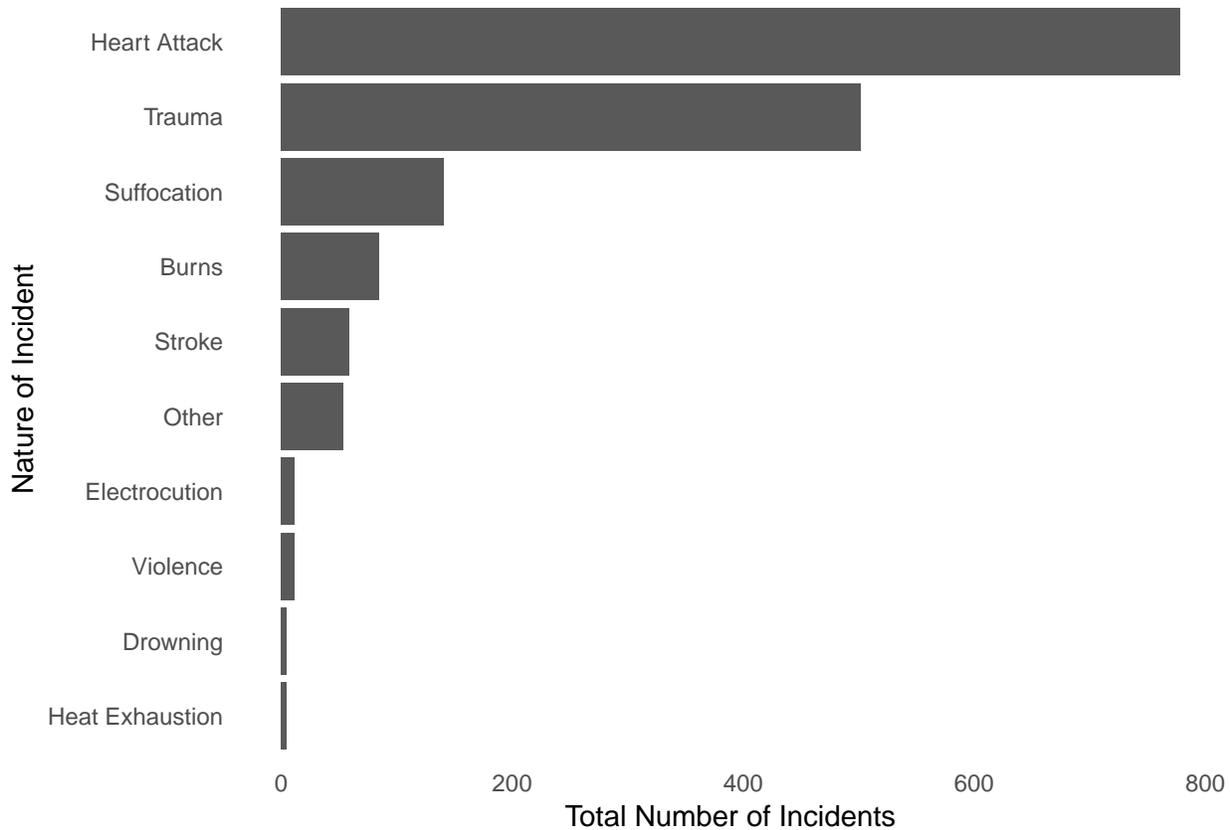


Figure 4: Line Chart

To get to this chart, I created the entire chunk of code I needed to get the visualization I wanted, and then simply replaced “Suffocation” in the filter function and ran it over and over again to see which Causes of Death had the largest fluctuation over time. Suffocation turned out to be the winner, so I chose to highlight it in this chart.

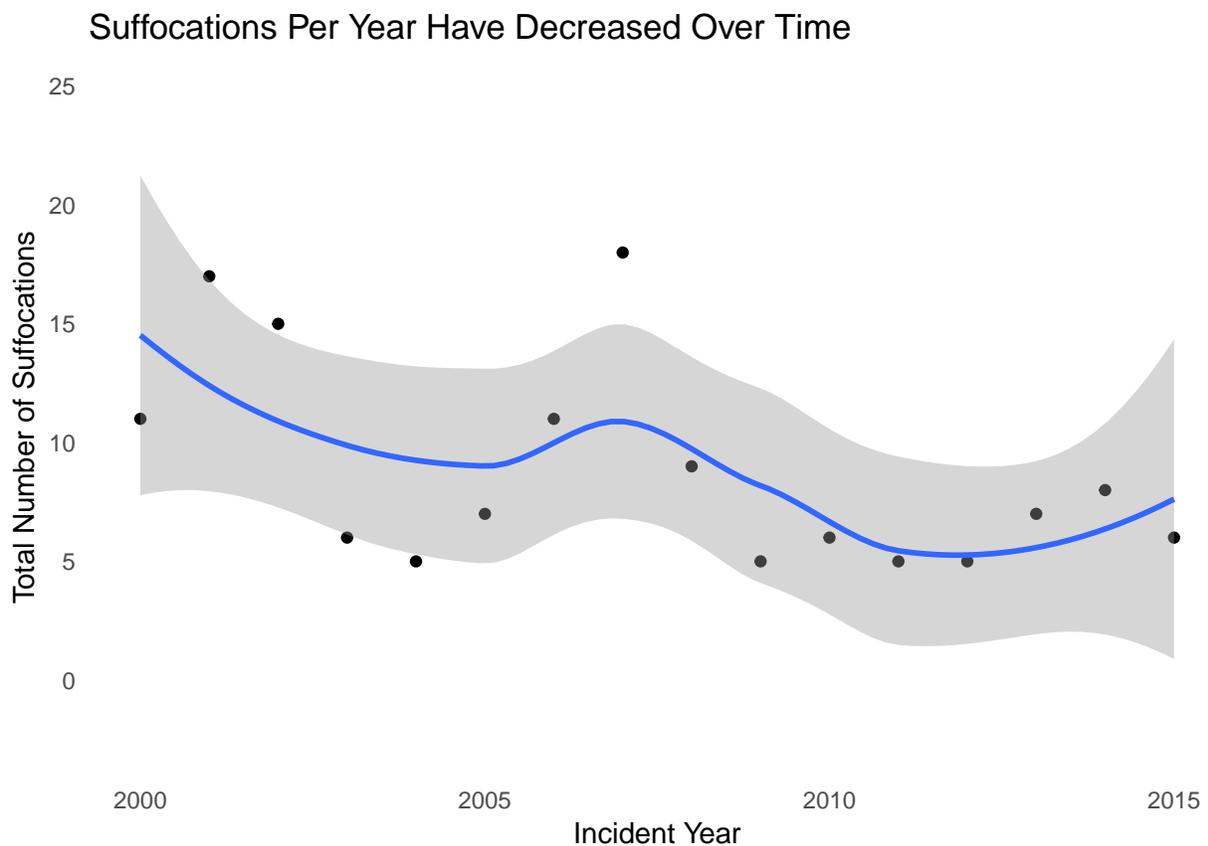
Again, I went with the Minimal theme and removed all gridlines, for consistency.

#What About Suffocation?

```
ff_suffocation <- Firefighters %>%
  mutate(Incident_Date = mdy(Incident_Date)) %>%
  mutate(Incident_Month = month(Incident_Date, label = TRUE)) %>%
  mutate(Incident_Year = year(Incident_Date)) %>%
  filter(Nature == "Suffocation") %>%
  filter(Incident_Year > 1999) %>%
  group_by(Incident_Year, Nature) %>%
  summarise(Total = n())

suffocations <- ggplot(ff_suffocation, aes(x = Incident_Year, y = Total)) +
  geom_point(stat = "identity") +
  labs(x= "Incident Year", y = "Total Number of Suffocations", title="Suffocations Per Year Have Decreased Over Time") +
  scale_y_continuous(expand = c(0, 5)) +
  stat_smooth() +
  theme_minimal() +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), legend.position = "bottom")

suffocations
```



Final Figure

This is the final figure, which combines the four charts together. There are some elements I chose to manipulate in Illustrator rather than R, like using the Proxima Nova font throughout for Titles, Axes, Body Copy, etc. I also changed some colors to be consistent with the legend (e.g. changing the standard error bar on the line chart from Gray to Green to indicate suffocation).

A lot of the design work I did was simply aligning different elements to one another (title to graphic, axis labels to one another, graphic sides to one another, etc). I also went in and added labels to call out specific areas of interest (e.g. 9/11, years with high/low numbers of suffocation deaths, etc). All of this was done in Illustrator.

I also added the source in the bottom right corner, and a disclaimer about filtering out 9/11 data to preserve the integrity of the data.

I got the “fire” color scheme from ColorBrewer online, and then created a Palette in R to apply to my different files. I then used that same Palette and colors to add highlights and flourishes to certain elements in Illustrator.

I believe that this conveys truth as best it can. I originally had this set up as four “quadrants,” but decided to order the four graphs on one line to facilitate the “story” I’m telling. I like the idea of zooming in and scrolling right to see each chart, rather than seeing all of them at once. This is a change I made at the last minute.

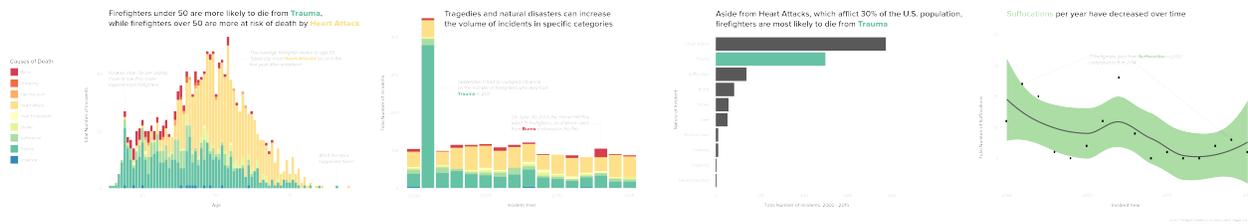


Figure 1: Firefighter Fatalities

Other Figures (Unused)

There were also a few charts I created while manipulating data and searching for the stories I wanted to tell. Below is my attempt to compare fatalities month-over-month, to see if any one particular month was more dangerous than others. As it turned out, the data was relatively consistent across months.

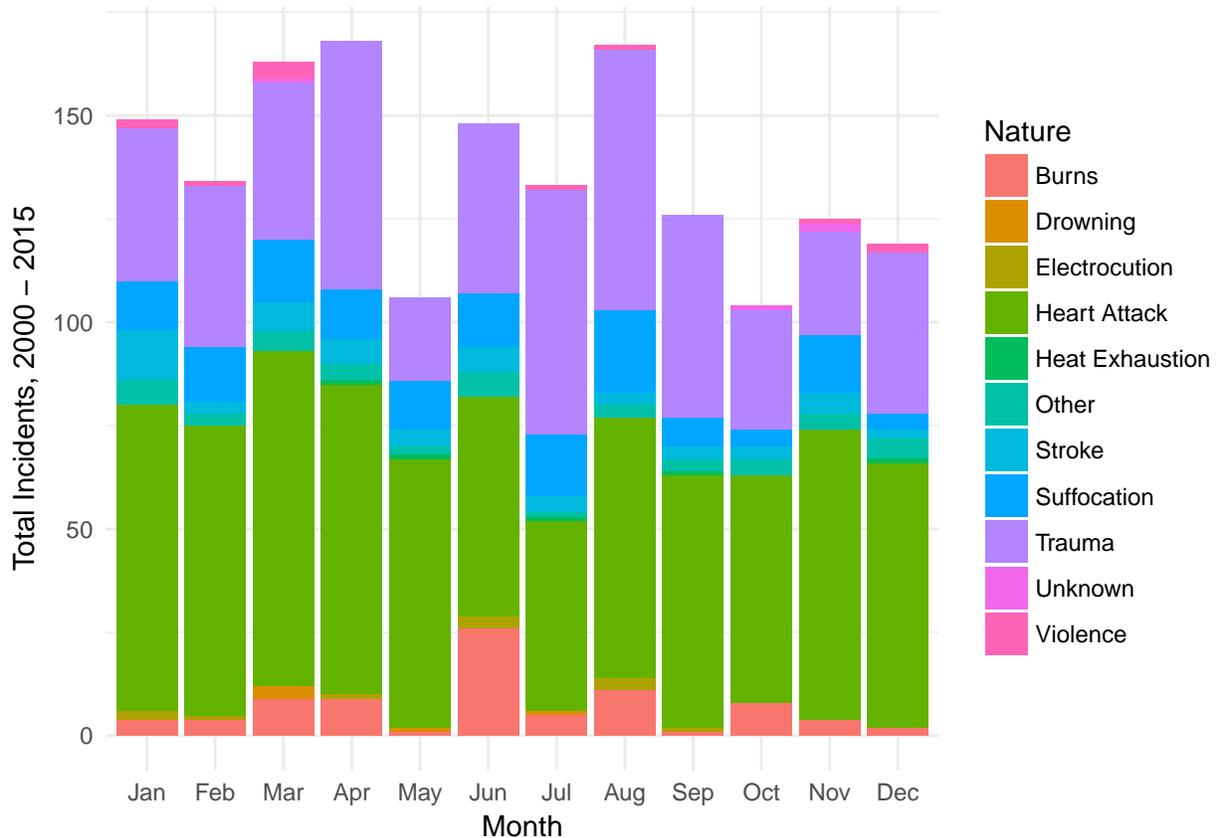
#Are Some Months More Dangerous Than Others?

```
ff_months <- Firefighters %>%
  mutate(Incident_Date = mdy(Incident_Date)) %>%
  mutate(Incident_Month = month(Incident_Date), label = TRUE) %>%
  mutate(Incident_Year = year(Incident_Date)) %>%
  filter(!Incident_Date == "2001-09-11") %>%
  filter(!Nature == "NA") %>%
  filter(Incident_Year > 1999) %>%
  group_by(Incident_Month, Nature) %>%
  summarise(Total = n())
```

```
by_month <- ggplot(ff_months, aes(x = Incident_Month, y = Total, fill = Nature, group = 1)) +
  geom_col() +
```

```
labs(x = "Month", y = "Total Incidents, 2000 - 2015") +
theme_minimal()
```

by_month



I also tried to break the data out by Season, which yielded similarly disappointing and confusing results.

#When Do Burn Deaths Tend to Occur?

```
ff_burns <- Firefighters %>%
  mutate(Incident_Date = mdy(Incident_Date)) %>%
  mutate(Incident_Month = month(Incident_Date, label = TRUE)) %>%
  mutate(Incident_Year = year(Incident_Date)) %>%
  filter(!Incident_Date == "2001-09-11") %>%
  filter(!Incident_Date == "2013-06-30") %>%
  filter(Nature == "Burns") %>%
  filter(Incident_Year > 1999) %>%
  mutate(Seasons = recode(Incident_Month,
    Jan = "Winter",
    Feb = "Winter",
    Mar = "Spring",
    Apr = "Spring",
    May = "Spring",
    Jun = "Summer",
    Jul = "Summer",
    Aug = "Summer",
    Sep = "Fall",
```

```

Oct = "Fall",
Nov = "Fall",
Dec = "Winter")) %>%
group_by(Incident_Year, Seasons, Nature) %>%
summarise(Total = n())

```

```

seasons <- ggplot(ff_burns, aes(x = Incident_Year, y = Total, color = Seasons)) +
geom_line(size = 1)+
facet_wrap(~ Seasons) +
theme_minimal()

```

seasons

