

2: Data

Empirical Methods

J. Seawright

Northwestern Political Science

Sept. 18, 2025

A perennially newsworthy question: how much political violence is there in the U.S., and what are its sources?

Let's look at the Global Terrorism Database.

In order to look at the data, I'm going to use a statistical analysis environment called R.

You can use R for free without installing anything at Posit Cloud.

```
head(usterrorism)
```

```
##          eventid iyear imonth iday approxdate extended resolution count:
## 1 197001010002  1970      1     1              0          20
## 2 197001020002  1970      1     2              0          20
## 3 197001020003  1970      1     2              0          20
## 4 197001030001  1970      1     3              0          20
## 5 197001050001  1970      1     1              0          20
## 6 197001060001  1970      1     6              0          20
##   country_txt region   region_txt provstate   city latitude longitude
## 1 United States      1 North America  Illinois   Cairo 37.00511  -89
## 2 United States      1 North America California Oakland 37.79193 -122
## 3 United States      1 North America Wisconsin Madison 43.07659  -89
## 4 United States      1 North America Wisconsin Madison 43.07295  -89
## 5 United States      1 North America Wisconsin Baraboo 43.46850  -89
## 6 United States      1 North America  Colorado  Denver 39.75897 -104
##   specificity vicinity      location
## 1           1          0
## 2           1          0 Edes Substation
## 3           1          0
```

```
table(usterrorism$ncill)
```

```
##  
##      0      1      2      3      4      5      6      7      8      9     10     11     13     15  
## 2726   224    52    10     8     4     2     2     3     1     3     2     1     2  
##    23    24    44    50    60   168   190  1385  
##     1     1     1     1     1     1     1     2
```

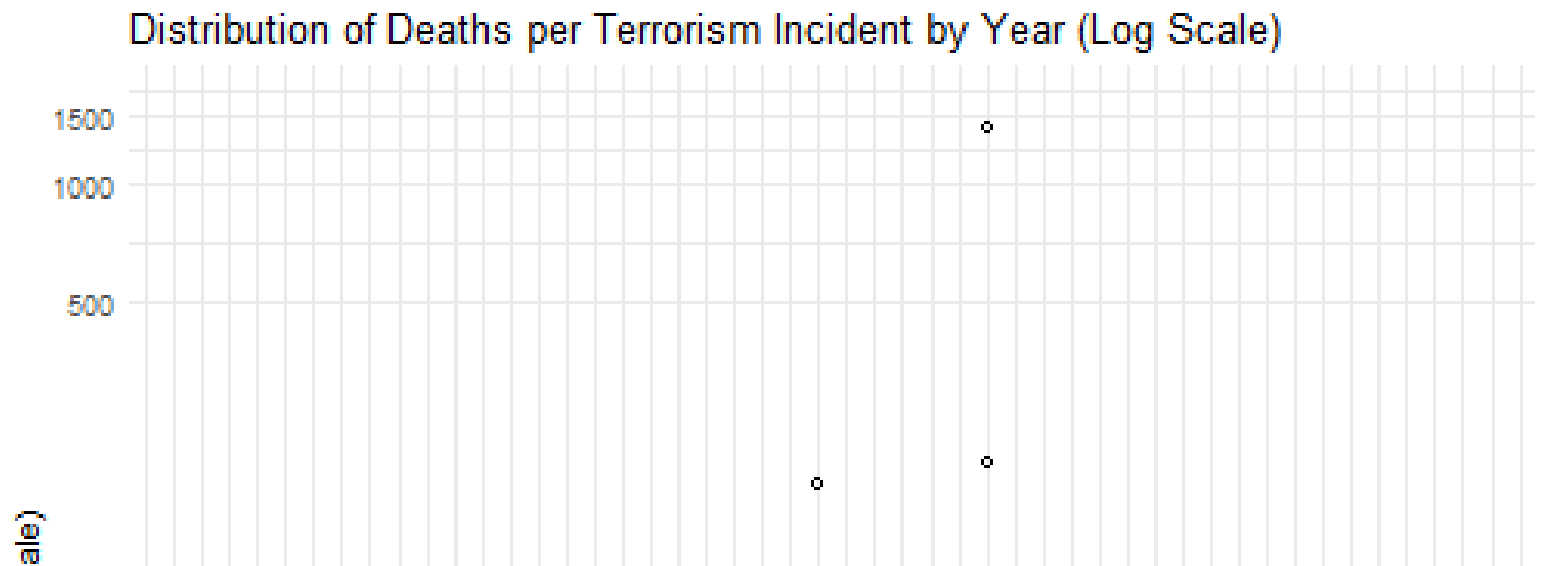
```
usterrorism %>% filter(nkill==1385)
```

```
##          eventid iyear imonth iday approxdate extended resolution count:
## 1 200109110004 2001      9    11              0          2:
## 2 200109110005 2001      9    11              0          2:
##      country_txt region      region_txt provstate      city latitude
## 1 United States      1 North America  New York New York City 40.69713
## 2 United States      1 North America  New York New York City 40.69713
##      specificity vicinity location
## 1              1          0
## 2              1          0
##
## 1
## 2 09/11/2001: This was one of four related attacks that occurred in tl
##      crit1 crit2 crit3 doubtterr alternative alternative_txt multiple suc
## 1      1      1      1          0          NA          1
## 2      1      1      1          0          NA          1
##      suicide attacktype1 attacktype1_txt attacktype2 attacktype2_txt att:
## 1      1          4      Hijacking          2      Armed Assault
## 2      1          4      Hijacking          2      Armed Assault
```

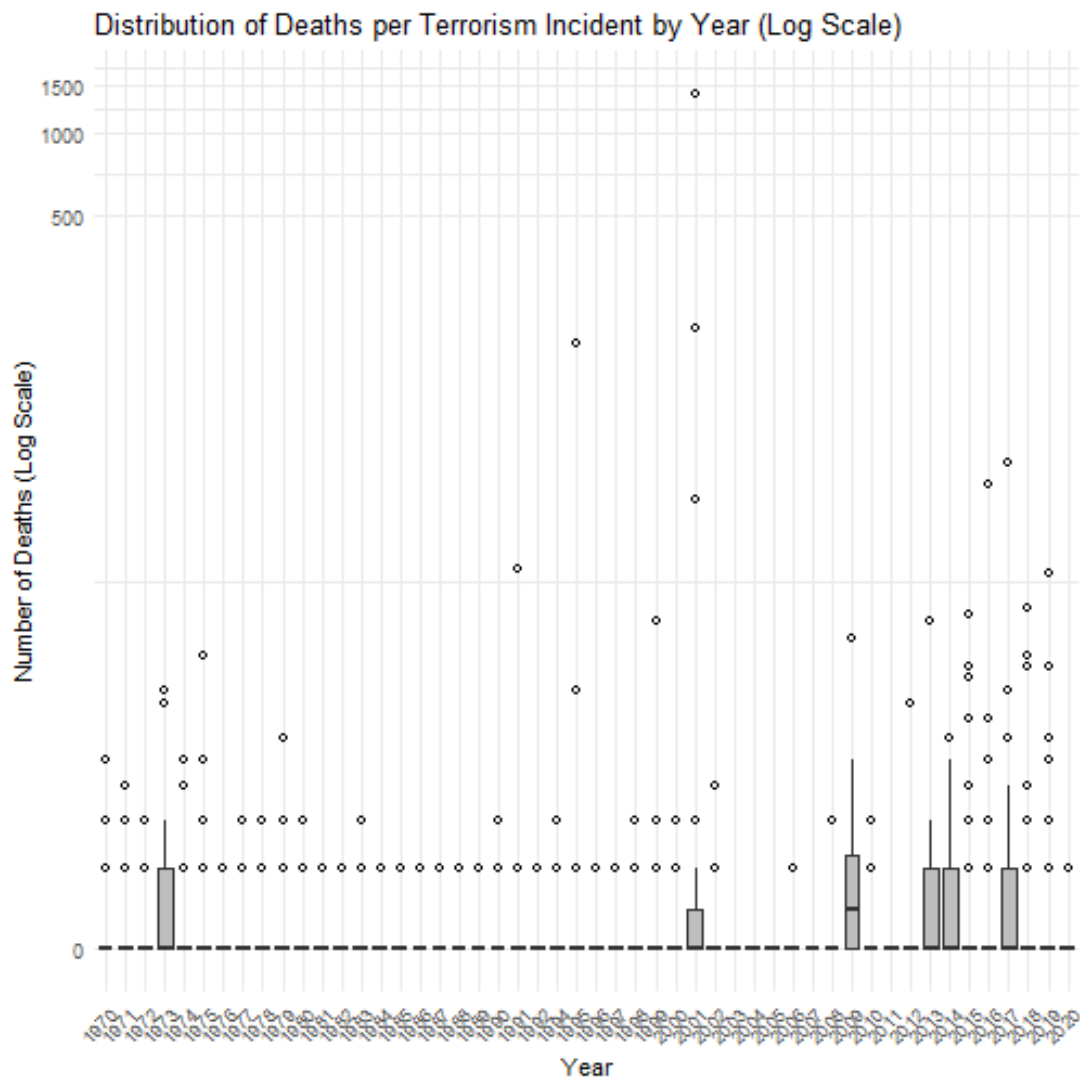


```
ggplot(usterrorism, aes(x = factor(iyear), y = nkill)) +
  geom_boxplot(fill = "grey", outlier.color = "black", outlier.shape = "circle",
    scale_y_continuous(trans = "log1p") + # This transforms the y-axis to a log scale
    labs(
      title = "Distribution of Deaths per Terrorism Incident by Year (Log Scale)",
      x = "Year",
      y = "Number of Deaths (Log Scale)"
    ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```
## Warning: Removed 70 rows containing non-finite outside the scale range [0, 1500].
## (`stat_boxplot()`.)
```



```
## Warning: Removed 70 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



```
table(usterrorism$gname)
```

Action Square

African-American extremists

Al-Qaida

Al-Qaida in the Arabian Peninsula (AQAP)

American Indian Movement

American Servicemen's Union (ASU)

Americans for a Competent Federal Judicial System

Americans for Justice

Anarchism

What is the Phineas Priesthood?

What about White Rabbit Three Percent Illinois
Patriot Freedom Fighters Militia?

Conclusions

1. Consider the perspective and limitations of data sources.
2. Quantitative data are often easier to work with than qualitative data, but may be less rewarding.
3. With enough work, it's usually possible to find or create data to learn an answer to your question.