```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.1     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.1     ✔ tibble    3.2.1
## ✔ lubridate 1.9.2     ✔ tidyr     1.3.0
## ✔ purrr     1.0.1
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflic
ts to become errors
```

```
raw_data <- read.csv("/Users/ibhargava/Desktop/globalterrorismdb_0718dist.csv")
data2 <- raw_data
```

```
summary(raw_data)
```

```
##     eventid              iyear          imonth            iday
##  Min.   :1.970e+11   Min.   :1970   Min.   : 0.000   Min.   : 0.00
##  1st Qu.:1.991e+11   1st Qu.:1991   1st Qu.: 4.000   1st Qu.: 8.00
##  Median :2.009e+11   Median :2009   Median : 6.000   Median :15.00
##  Mean   :2.003e+11   Mean   :2003   Mean   : 6.467   Mean   :15.51
##  3rd Qu.:2.014e+11   3rd Qu.:2014   3rd Qu.: 9.000   3rd Qu.:23.00
##  Max.   :2.017e+11   Max.   :2017   Max.   :12.000   Max.   :31.00
##
##   approxdate           extended        resolution           country
##  Length:181691      Min.   :0.00000   Length:181691      Min.   :   4
##  Class :character   1st Qu.:0.00000   Class :character   1st Qu.:  78
##  Mode  :character   Median :0.00000   Mode  :character   Median :  98
##                     Mean   :0.04535                      Mean   : 132
##                     3rd Qu.:0.00000                      3rd Qu.: 160
##                     Max.   :1.00000                      Max.   :1004
##
##  country_txt            region        region_txt         provstate
##  Length:181691      Min.   : 1.000   Length:181691      Length:181691
##  Class :character   1st Qu.: 5.000   Class :character   Class :character
##  Mode  :character   Median : 6.000   Mode  :character   Mode  :character
##                     Mean   : 7.161
##                     3rd Qu.:10.000
```

```
##                   Max.   :12.000
##
##      city              latitude        longitude          specificity
##  Length:181691      Min.   :-53.16   Min.   :-86185896   Min.   :1.000
##  Class :character   1st Qu.: 11.51   1st Qu.:        5   1st Qu.:1.000
##  Mode  :character   Median : 31.47   Median :       43   Median :1.000
##                     Mean   : 23.50   Mean   :     -459   Mean   :1.451
##                     3rd Qu.: 34.69   3rd Qu.:       69   3rd Qu.:1.000
##                     Max.   : 74.63   Max.   :      179   Max.   :5.000
##                     NA's   :4556     NA's   :4557        NA's   :6
##     vicinity          location           summary             crit1
##  Min.   :-9.0000   Length:181691      Length:181691      Min.   :0.0000
##  1st Qu.: 0.0000   Class :character   Class :character   1st Qu.:1.0000
##  Median : 0.0000   Mode  :character   Mode  :character   Median :1.0000
##  Mean   : 0.0683                                        Mean   :0.9885
##  3rd Qu.: 0.0000                                        3rd Qu.:1.0000
##  Max.   : 1.0000                                        Max.   :1.0000
##
##      crit2             crit3          doubtterr        alternative
##  Min.   :0.0000   Min.   :0.0000   Min.   :-9.0000   Min.   :1.00
##  1st Qu.:1.0000   1st Qu.:1.0000   1st Qu.: 0.0000   1st Qu.:1.00
##  Median :1.0000   Median :1.0000   Median : 0.0000   Median :1.00
##  Mean   :0.9931   Mean   :0.8757   Mean   :-0.5232   Mean   :1.29
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.: 0.0000   3rd Qu.:1.00
##  Max.   :1.0000   Max.   :1.0000   Max.   : 1.0000   Max.   :5.00
##                                    NA's   :1         NA's   :152680
##  alternative_txt       multiple         success          suicide
##  Length:181691      Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##  Class :character   1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:0.00000
##  Mode  :character   Median :0.0000   Median :1.0000   Median :0.00000
##                     Mean   :0.1378   Mean   :0.8896   Mean   :0.03651
##                     3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:0.00000
##                     Max.   :1.0000   Max.   :1.0000   Max.   :1.00000
##                     NA's   :1
##   attacktype1    attacktype1_txt     attacktype2     attacktype2_txt
##  Min.   :1.000   Length:181691      Min.   :1.00     Length:181691
##  1st Qu.:2.000   Class :character   1st Qu.:2.00     Class :character
##  Median :3.000   Mode  :character   Median :2.00     Mode  :character
##  Mean   :3.248                      Mean   :3.72
##  3rd Qu.:3.000                      3rd Qu.:7.00
##  Max.   :9.000                      Max.   :9.00
##                                     NA's   :175377
##   attacktype3     attacktype3_txt      targtype1     targtype1_txt
##  Min.   :1.00     Length:181691      Min.   : 1.00   Length:181691
##  1st Qu.:2.00     Class :character   1st Qu.: 3.00   Class :character
##  Median :7.00     Mode  :character   Median : 4.00   Mode  :character
##  Mean   :5.25                        Mean   : 8.44
```

```
##  3rd Qu.:7.00                       3rd Qu.:14.00
##  Max.   :8.00                       Max.   :22.00
##  NA's   :181263
##   targsubtype1    targsubtype1_txt      corp1             target1
##  Min.   :  1.00   Length:181691      Length:181691      Length:181691
##  1st Qu.: 22.00   Class :character   Class :character   Class :character
##  Median : 35.00   Mode  :character   Mode  :character   Mode  :character
##  Mean   : 46.97
##  3rd Qu.: 74.00
##  Max.   :113.00
##  NA's   :10373
##     natlty1       natlty1_txt          targtype2      targtype2_txt
##  Min.   :   4.0   Length:181691      Min.   : 1.00    Length:181691
##  1st Qu.:  83.0   Class :character   1st Qu.: 4.00    Class :character
##  Median : 101.0   Mode  :character   Median :14.00    Mode  :character
##  Mean   : 127.7                      Mean   :10.25
##  3rd Qu.: 173.0                      3rd Qu.:14.00
##  Max.   :1004.0                      Max.   :22.00
##  NA's   :1559                        NA's   :170547
##   targsubtype2    targsubtype2_txt      corp2             target2
##  Min.   :  1.00   Length:181691      Length:181691      Length:181691
##  1st Qu.: 34.00   Class :character   Class :character   Class :character
##  Median : 67.00   Mode  :character   Mode  :character   Mode  :character
##  Mean   : 55.31
##  3rd Qu.: 69.00
##  Max.   :113.00
##  NA's   :171006
##     natlty2       natlty2_txt          targtype3      targtype3_txt
##  Min.   :   4.0   Length:181691      Min.   : 1.00    Length:181691
##  1st Qu.:  92.0   Class :character   1st Qu.: 3.00    Class :character
##  Median :  98.0   Mode  :character   Median :14.00    Mode  :character
##  Mean   : 131.2                      Mean   :10.02
##  3rd Qu.: 182.0                      3rd Qu.:14.00
##  Max.   :1004.0                      Max.   :22.00
##  NA's   :170863                      NA's   :180515
##   targsubtype3    targsubtype3_txt      corp3             target3
##  Min.   :  1.00   Length:181691      Length:181691      Length:181691
##  1st Qu.: 33.00   Class :character   Class :character   Class :character
##  Median : 67.00   Mode  :character   Mode  :character   Mode  :character
##  Mean   : 55.55
##  3rd Qu.: 73.00
##  Max.   :113.00
##  NA's   :180594
##     natlty3       natlty3_txt          gname             gsubname
##  Min.   :   4.0   Length:181691      Length:181691      Length:181691
##  1st Qu.:  75.0   Class :character   Class :character   Class :character
##  Median : 110.0   Mode  :character   Mode  :character   Mode  :character
```

```
##  Mean   : 144.6
##  3rd Qu.: 182.0
##  Max.   :1004.0
##  NA's   :180544
##     gname2           gsubname2            gname3           gsubname3
##  Length:181691      Length:181691      Length:181691      Length:181691
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##     motive           guncertain1      guncertain2      guncertain3
##  Length:181691      Min.   :0.0000   Min.   :0.00     Min.   :0.00
##  Class :character   1st Qu.:0.0000   1st Qu.:0.00     1st Qu.:0.00
##  Mode  :character   Median :0.0000   Median :0.00     Median :0.00
##                     Mean   :0.0814   Mean   :0.27     Mean   :0.19
##                     3rd Qu.:0.0000   3rd Qu.:1.00     3rd Qu.:0.00
##                     Max.   :1.0000   Max.   :1.00     Max.   :1.00
##                     NA's   :380      NA's   :179736   NA's   :181371
##    individual          nperps            nperpcap         claimed
##  Min.   :0.00000   Min.   :  -99.00   Min.   :-99.00   Min.   :-9.00
##  1st Qu.:0.00000   1st Qu.:  -99.00   1st Qu.:  0.00   1st Qu.: 0.00
##  Median :0.00000   Median :  -99.00   Median :  0.00   Median : 0.00
##  Mean   :0.00295   Mean   :  -65.36   Mean   : -1.52   Mean   : 0.05
##  3rd Qu.:0.00000   3rd Qu.:    1.00   3rd Qu.:  0.00   3rd Qu.: 0.00
##  Max.   :1.00000   Max.   :25000.00   Max.   :406.00   Max.   : 1.00
##                    NA's   :71115      NA's   :69489    NA's   :66120
##    claimmode      claimmode_txt          claim2          claimmode2
##  Min.   : 1.00    Length:181691      Min.   :-9.00    Min.   : 1.00
##  1st Qu.: 6.00    Class :character   1st Qu.: 0.00    1st Qu.: 6.00
##  Median : 8.00    Mode  :character   Median : 0.00    Median : 7.00
##  Mean   : 7.02                       Mean   : 0.25    Mean   : 7.18
##  3rd Qu.: 8.00                       3rd Qu.: 1.00    3rd Qu.:10.00
##  Max.   :10.00                       Max.   : 1.00    Max.   :10.00
##  NA's   :162608                      NA's   :179801   NA's   :181075
##  claimmode2_txt         claim3          claimmode3     claimmode3_txt
##  Length:181691      Min.   :0.00     Min.   : 1.00    Length:181691
##  Class :character   1st Qu.:0.00     1st Qu.: 4.00    Class :character
##  Mode  :character   Median :0.00     Median : 7.00    Mode  :character
##                     Mean   :0.41     Mean   : 6.73
##                     3rd Qu.:1.00     3rd Qu.: 9.00
##                     Max.   :1.00     Max.   :10.00
##                     NA's   :181373   NA's   :181558
##    compclaim         weaptype1      weaptype1_txt       weapsubtype1
##  Min.   :-9.0     Min.   : 1.000   Length:181691      Min.   : 1.00
##  1st Qu.:-9.0     1st Qu.: 5.000   Class :character   1st Qu.: 5.00
```

```
##  Median :-9.0     Median : 6.000   Mode  :character   Median :12.00
##  Mean   :-6.3     Mean   : 6.447                      Mean   :11.12
##  3rd Qu.: 0.0     3rd Qu.: 6.000                      3rd Qu.:16.00
##  Max.   : 1.0     Max.   :13.000                      Max.   :31.00
##  NA's   :176852                                       NA's   :20768
##  weapsubtype1_txt     weaptype2      weaptype2_txt       weapsubtype2
##  Length:181691      Min.   : 1.00    Length:181691      Min.   : 1.00
##  Class :character   1st Qu.: 5.00    Class :character   1st Qu.: 5.00
##  Mode  :character   Median : 6.00    Mode  :character   Median : 7.00
##                     Mean   : 6.81                       Mean   :10.75
##                     3rd Qu.: 8.00                       3rd Qu.:18.00
##                     Max.   :13.00                       Max.   :31.00
##                     NA's   :168564                      NA's   :170149
##  weapsubtype2_txt     weaptype3      weaptype3_txt       weapsubtype3
##  Length:181691      Min.   : 2.00    Length:181691      Min.   : 1.00
##  Class :character   1st Qu.: 5.00    Class :character   1st Qu.: 4.00
##  Mode  :character   Median : 6.00    Mode  :character   Median : 7.00
##                     Mean   : 6.91                       Mean   :11.64
##                     3rd Qu.: 9.00                       3rd Qu.:20.00
##                     Max.   :13.00                       Max.   :28.00
##                     NA's   :179828                      NA's   :179998
##  weapsubtype3_txt     weaptype4      weaptype4_txt       weapsubtype4
##  Length:181691      Min.   : 5.00    Length:181691      Min.   : 2.00
##  Class :character   1st Qu.: 5.00    Class :character   1st Qu.: 3.00
##  Mode  :character   Median : 6.00    Mode  :character   Median : 9.50
##                     Mean   : 6.25                       Mean   :10.84
##                     3rd Qu.: 6.00                       3rd Qu.:16.00
##                     Max.   :12.00                       Max.   :28.00
##                     NA's   :181618                      NA's   :181621
##  weapsubtype4_txt    weapdetail            nkill             nkillus
##  Length:181691      Length:181691      Min.   :   0.000   Min.   :   0.00
##  Class :character   Class :character   1st Qu.:   0.000   1st Qu.:   0.00
##  Mode  :character   Mode  :character   Median :   0.000   Median :   0.00
##                                        Mean   :   2.403   Mean   :   0.05
##                                        3rd Qu.:   2.000   3rd Qu.:   0.00
##                                        Max.   :1570.000   Max.   :1360.00
##                                        NA's   :10313      NA's   :64446
##     nkillter          nwound            nwoundus         nwoundte
##  Min.   :  0.00   Min.   :   0.000   Min.   :  0.00   Min.   :  0.00
##  1st Qu.:  0.00   1st Qu.:   0.000   1st Qu.:  0.00   1st Qu.:  0.00
##  Median :  0.00   Median :   0.000   Median :  0.00   Median :  0.00
##  Mean   :  0.51   Mean   :   3.168   Mean   :  0.04   Mean   :  0.11
##  3rd Qu.:  0.00   3rd Qu.:   2.000   3rd Qu.:  0.00   3rd Qu.:  0.00
##  Max.   :500.00   Max.   :8191.000   Max.   :751.00   Max.   :200.00
##  NA's   :66958    NA's   :16311      NA's   :64702    NA's   :69143
##     property         propextent     propextent_txt       propvalue
##  Min.   :-9.0000   Min.   :1.0      Length:181691      Min.   :       -99
```

```
##  1st Qu.: 0.0000   1st Qu.:3.0      Class :character   1st Qu.:       -99
##  Median : 1.0000   Median :3.0      Mode  :character   Median :       -99
##  Mean   :-0.5446   Mean   :3.3                         Mean   :    208812
##  3rd Qu.: 1.0000   3rd Qu.:4.0                         3rd Qu.:      1000
##  Max.   : 1.0000   Max.   :4.0                         Max.   :2700000000
##                    NA's   :117626                      NA's   :142702
##  propcomment          ishostkid           nhostkid          nhostkidus
##  Length:181691      Min.   :-9.00000   Min.   :  -99.00   Min.   :-99.00
##  Class :character   1st Qu.: 0.00000   1st Qu.:    1.00   1st Qu.:  0.00
##  Mode  :character   Median : 0.00000   Median :    2.00   Median :  0.00
##                     Mean   : 0.05905   Mean   :    4.53   Mean   : -0.35
##                     3rd Qu.: 0.00000   3rd Qu.:    4.00   3rd Qu.:  0.00
##                     Max.   : 1.00000   Max.   :17000.00   Max.   : 86.00
##                     NA's   :178        NA's   :168119     NA's   :168174
##      nhours            ndays            divert          kidhijcountry
##  Min.   :-99.00   Min.   : -99.00   Length:181691      Length:181691
##  1st Qu.:-99.00   1st Qu.: -99.00   Class :character   Class :character
##  Median :-99.00   Median : -99.00   Mode  :character   Mode  :character
##  Mean   :-46.79   Mean   : -32.52
##  3rd Qu.:  0.00   3rd Qu.:   4.00
##  Max.   :999.00   Max.   :2454.00
##  NA's   :177628   NA's   :173567
##      ransom          ransomamt            ransomamtus           ransompaid
##  Min.   :-9.00    Min.   :       -99   Min.   :      -99   Min.   :      -99
##  1st Qu.: 0.00    1st Qu.:         0   1st Qu.:        0   1st Qu.:      -99
##  Median : 0.00    Median :     15000   Median :        0   Median :        0
##  Mean   :-0.15    Mean   :   3172530   Mean   :   578487   Mean   :   717944
##  3rd Qu.: 0.00    3rd Qu.:    400000   3rd Qu.:        0   3rd Qu.:     1273
##  Max.   : 1.00    Max.   :1000000000   Max.   :132000000   Max.   :275000000
##  NA's   :104310   NA's   :180341       NA's   :181128      NA's   :180917
##   ransompaidus      ransomnote        hostkidoutcome   hostkidoutcome_txt
##  Min.   :  -99.0   Length:181691      Min.   :1.00     Length:181691
##  1st Qu.:    0.0   Class :character   1st Qu.:2.00     Class :character
##  Median :    0.0   Mode  :character   Median :4.00     Mode  :character
##  Mean   :  240.4                      Mean   :4.63
##  3rd Qu.:    0.0                      3rd Qu.:7.00
##  Max.   :48000.0                      Max.   :7.00
##  NA's   :181139                       NA's   :170700
##    nreleased         addnotes            scite1             scite2
##  Min.   : -99.00   Length:181691      Length:181691      Length:181691
##  1st Qu.: -99.00   Class :character   Class :character   Class :character
##  Median :   0.00   Mode  :character   Mode  :character   Mode  :character
##  Mean   : -29.02
##  3rd Qu.:   1.00
##  Max.   :2769.00
##  NA's   :171291
##     scite3            dbsource            INT_LOG          INT_IDEO
```

```
##  Length:181691      Length:181691      Min.   :-9.000   Min.   :-9.000
##  Class :character   Class :character   1st Qu.:-9.000   1st Qu.:-9.000
##  Mode  :character   Mode  :character   Median :-9.000   Median :-9.000
##                                        Mean   :-4.544   Mean   :-4.464
##                                        3rd Qu.: 0.000   3rd Qu.: 0.000
##                                        Max.   : 1.000   Max.   : 1.000
##
##     INT_MISC           INT_ANY         related
##  Min.   :-9.00000   Min.   :-9.000   Length:181691
##  1st Qu.: 0.00000   1st Qu.:-9.000   Class :character
##  Median : 0.00000   Median : 0.000   Mode  :character
##  Mean   : 0.09001   Mean   :-3.946
##  3rd Qu.: 0.00000   3rd Qu.: 0.000
##  Max.   : 1.00000   Max.   : 1.000
##
```

```
sapply(raw_data, function(x) sum(is.na(x)))
```

```
##            eventid              iyear             imonth               iday
##                  0                  0                  0                  0
##         approxdate           extended         resolution            country
##                  0                  0                  0                  0
##       country_txt             region        region_txt          provstate
##                  0                  0                  0                  0
##               city           latitude          longitude        specificity
##                  0               4556               4557                  6
##           vicinity           location            summary              crit1
##                  0                  0                  0                  0
##              crit2              crit3         doubtterr        alternative
##                  0                  0                  1             152680
##    alternative_txt           multiple            success            suicide
##                  0                  1                  0                  0
##        attacktype1    attacktype1_txt        attacktype2    attacktype2_txt
##                  0                  0             175377                  0
##        attacktype3    attacktype3_txt         targtype1      targtype1_txt
##             181263                  0                  0                  0
##       targsubtype1   targsubtype1_txt              corp1            target1
##              10373                  0                 33                  1
##            natlty1        natlty1_txt         targtype2      targtype2_txt
##               1559                  0             170547                  0
##       targsubtype2   targsubtype2_txt              corp2            target2
##             171006                  0                  0                  0
##            natlty2        natlty2_txt         targtype3      targtype3_txt
##             170863                  0             180515                  0
##       targsubtype3   targsubtype3_txt              corp3            target3
##             180594                  0                  0                  0
```

```
##            natlty3        natlty3_txt              gname           gsubname
##             180544                  0                  0                  0
##             gname2         gsubname2             gname3          gsubname3
##                  0                  0                  0                  0
##             motive        guncertain1        guncertain2        guncertain3
##                  0                380             179736             181371
##         individual             nperps           nperpcap            claimed
##                  0              71115              69489              66120
##          claimmode      claimmode_txt             claim2         claimmode2
##             162608                  0             179801             181075
##     claimmode2_txt             claim3         claimmode3     claimmode3_txt
##                  0             181373             181558                  0
##          compclaim          weaptype1      weaptype1_txt       weapsubtype1
##             176852                  0                  0              20768
##   weapsubtype1_txt          weaptype2      weaptype2_txt       weapsubtype2
##                  0             168564                  0             170149
##   weapsubtype2_txt          weaptype3      weaptype3_txt       weapsubtype3
##                  0             179828                  0             179998
##   weapsubtype3_txt          weaptype4      weaptype4_txt       weapsubtype4
##                  0             181618                  0             181621
##   weapsubtype4_txt         weapdetail              nkill            nkillus
##                  0                  0              10313              64446
##           nkillter             nwound           nwoundus           nwoundte
##              66958              16311              64702              69143
##           property         propextent     propextent_txt           propvalue
##                  0             117626                  0             142702
##        propcomment          ishostkid          nhostkid          nhostkidus
##                  0                178             168119             168174
##             nhours              ndays             divert      kidhijcountry
##             177628             173567                  0                  0
##             ransom          ransomamt        ransomamtus         ransompaid
##             104310             180341             181128             180917
##       ransompaidus        ransomnote     hostkidoutcome hostkidoutcome_txt
##             181139                  0             170700                  0
##          nreleased           addnotes             scite1             scite2
##             171291                  0                  0                  0
##             scite3           dbsource            INT_LOG            INT_IDEO
##                  0                  0                  0                  0
##           INT_MISC            INT_ANY            related
##                  0                  0                  0
```

```
data2[is.na(data2)] <- 0
```

```
sapply(data2, function(x) sum(is.na(x)))
```

```
##            eventid              iyear             imonth               iday
##                  0                  0                  0                  0
##         approxdate           extended         resolution            country
##                  0                  0                  0                  0
##        country_txt             region         region_txt          provstate
##                  0                  0                  0                  0
##               city           latitude          longitude        specificity
##                  0                  0                  0                  0
##           vicinity           location            summary              crit1
##                  0                  0                  0                  0
##              crit2              crit3          doubtterr        alternative
##                  0                  0                  0                  0
##    alternative_txt           multiple            success            suicide
##                  0                  0                  0                  0
##        attacktype1    attacktype1_txt        attacktype2    attacktype2_txt
##                  0                  0                  0                  0
##        attacktype3    attacktype3_txt         targtype1      targtype1_txt
##                  0                  0                  0                  0
##       targsubtype1   targsubtype1_txt              corp1            target1
##                  0                  0                  0                  0
##            natlty1        natlty1_txt         targtype2      targtype2_txt
##                  0                  0                  0                  0
##       targsubtype2   targsubtype2_txt              corp2            target2
##                  0                  0                  0                  0
##            natlty2        natlty2_txt         targtype3      targtype3_txt
##                  0                  0                  0                  0
##       targsubtype3   targsubtype3_txt              corp3            target3
##                  0                  0                  0                  0
##            natlty3        natlty3_txt              gname           gsubname
##                  0                  0                  0                  0
##             gname2          gsubname2             gname3          gsubname3
##                  0                  0                  0                  0
##             motive        guncertain1        guncertain2        guncertain3
##                  0                  0                  0                  0
##         individual             nperps           nperpcap            claimed
##                  0                  0                  0                  0
##          claimmode      claimmode_txt             claim2         claimmode2
##                  0                  0                  0                  0
##     claimmode2_txt             claim3         claimmode3     claimmode3_txt
##                  0                  0                  0                  0
##          compclaim          weaptype1      weaptype1_txt       weapsubtype1
##                  0                  0                  0                  0
##    weapsubtype1_txt          weaptype2      weaptype2_txt       weapsubtype2
##                  0                  0                  0                  0
##    weapsubtype2_txt          weaptype3      weaptype3_txt       weapsubtype3
##                  0                  0                  0                  0
##    weapsubtype3_txt          weaptype4      weaptype4_txt       weapsubtype4
```

```
##                   0                   0                   0                   0
##   weapsubtype4_txt         weapdetail               nkill            nkillus
##                   0                   0                   0                   0
##           nkillter              nwound           nwoundus           nwoundte
##                   0                   0                   0                   0
##           property          propextent     propextent_txt           propvalue
##                   0                   0                   0                   0
##        propcomment           ishostkid           nhostkid         nhostkidus
##                   0                   0                   0                   0
##              nhours               ndays              divert      kidhijcountry
##                   0                   0                   0                   0
##              ransom           ransomamt        ransomamtus         ransompaid
##                   0                   0                   0                   0
##       ransompaidus         ransomnote     hostkidoutcome hostkidoutcome_txt
##                   0                   0                   0                   0
##          nreleased            addnotes              scite1              scite2
##                   0                   0                   0                   0
##              scite3            dbsource             INT_LOG            INT_IDEO
##                   0                   0                   0                   0
##            INT_MISC             INT_ANY             related
##                   0                   0                   0
```

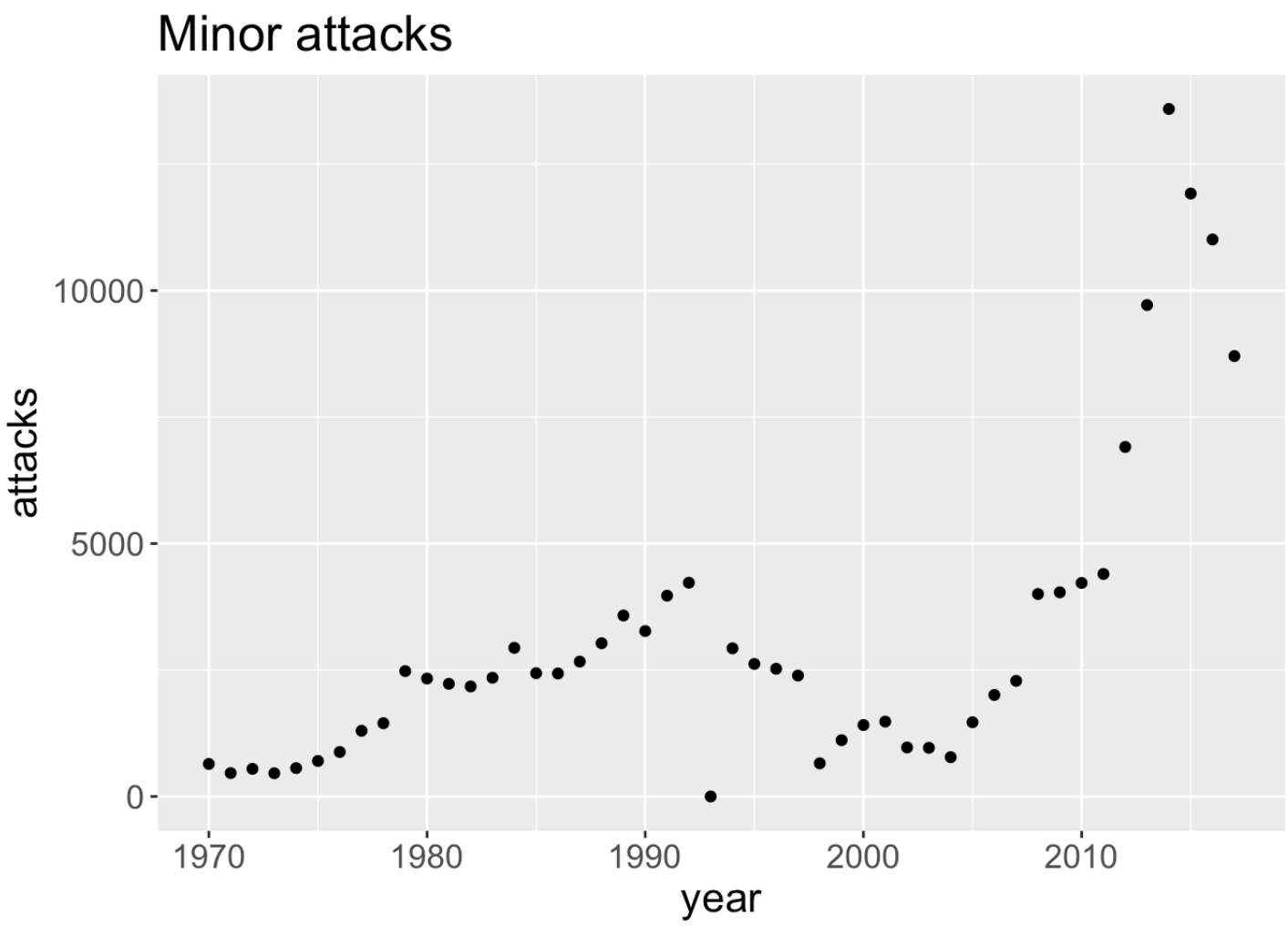
```
data2$attack_var <- data2$nkill
```

```
data2 <- data2 %>% mutate(attack_var = replace(attack_var,data2$nkill>=3& data2$nkill
<10,2))
data2 <- data2 %>% mutate(attack_var = replace(attack_var, data2$nkill>=10,3))
data2 <- data2 %>% mutate(attack_var = replace(attack_var, data2$nkill<3,1))
```

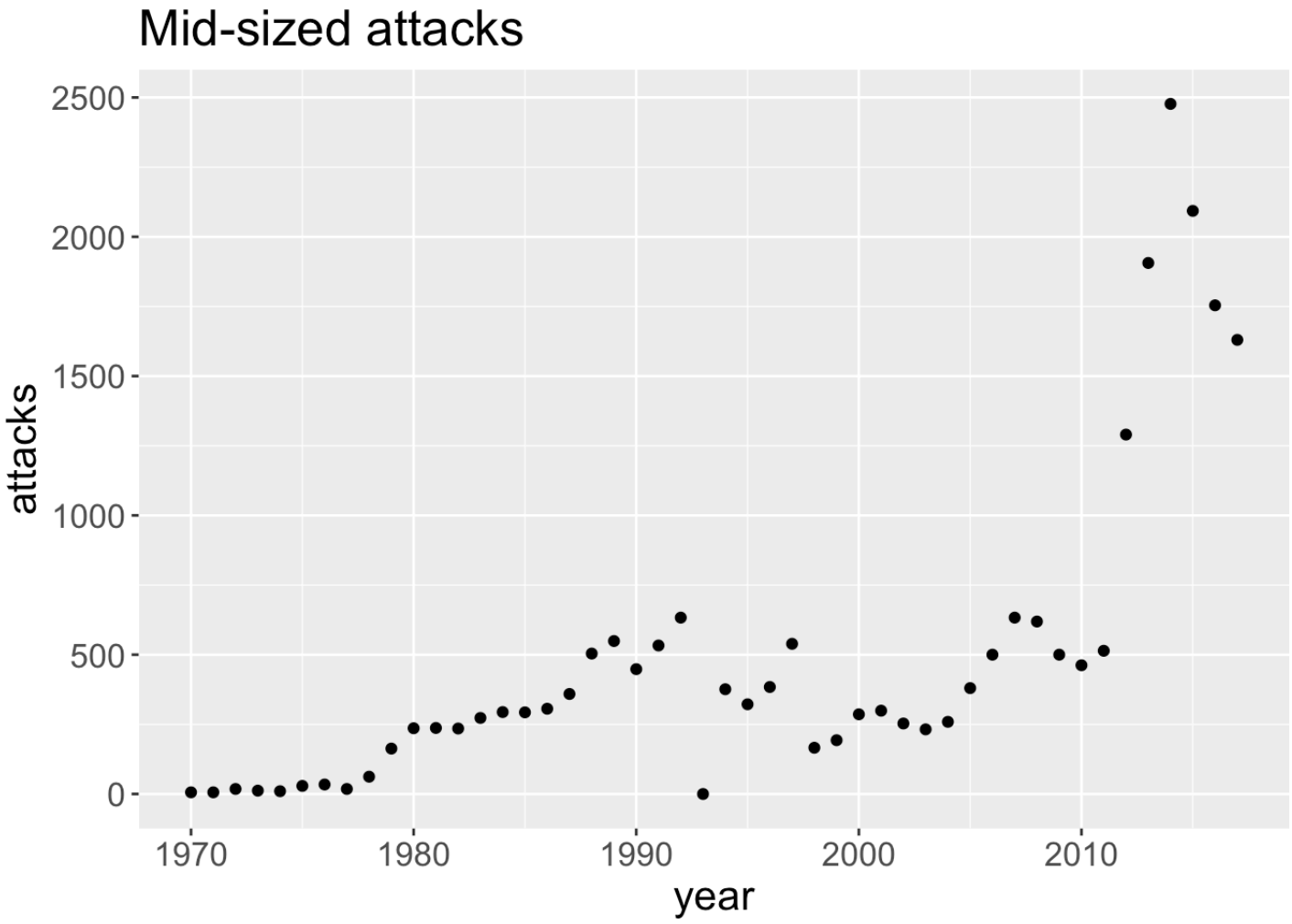
```
df1 <- data2[data2$attack_var == "1", ]
df2<- data2[data2$attack_var == "2", ]
df3<-data2[data2$attack_var == "3", ]

df1_num = data.frame(year=1970,df1 %>% filter(df1$iyear==1970) %>% summarize(attacks
= n()))
df2_num = data.frame(year=1970,df2 %>% filter(df2$iyear==1970) %>% summarize(attacks
= n()))
df3_num = data.frame(year=1970,df3 %>% filter(df3$iyear==1970) %>% summarize(attacks
= n()))

for (n in 1971:2017)
{
    df1_num[nrow(df1_num) + 1,] = c(n, df1 %>% filter(df1$iyear==n) %>% summarize(att
acks = n()))
    df2_num[nrow(df2_num) + 1,] = c(n, df2 %>% filter(df2$iyear==n) %>% summarize(att
acks = n()))
    df3_num[nrow(df3_num) + 1,] = c(n, df3 %>% filter(df3$iyear==n) %>% summarize(att
acks = n()))
}
```

```
library(ggplot2)
```

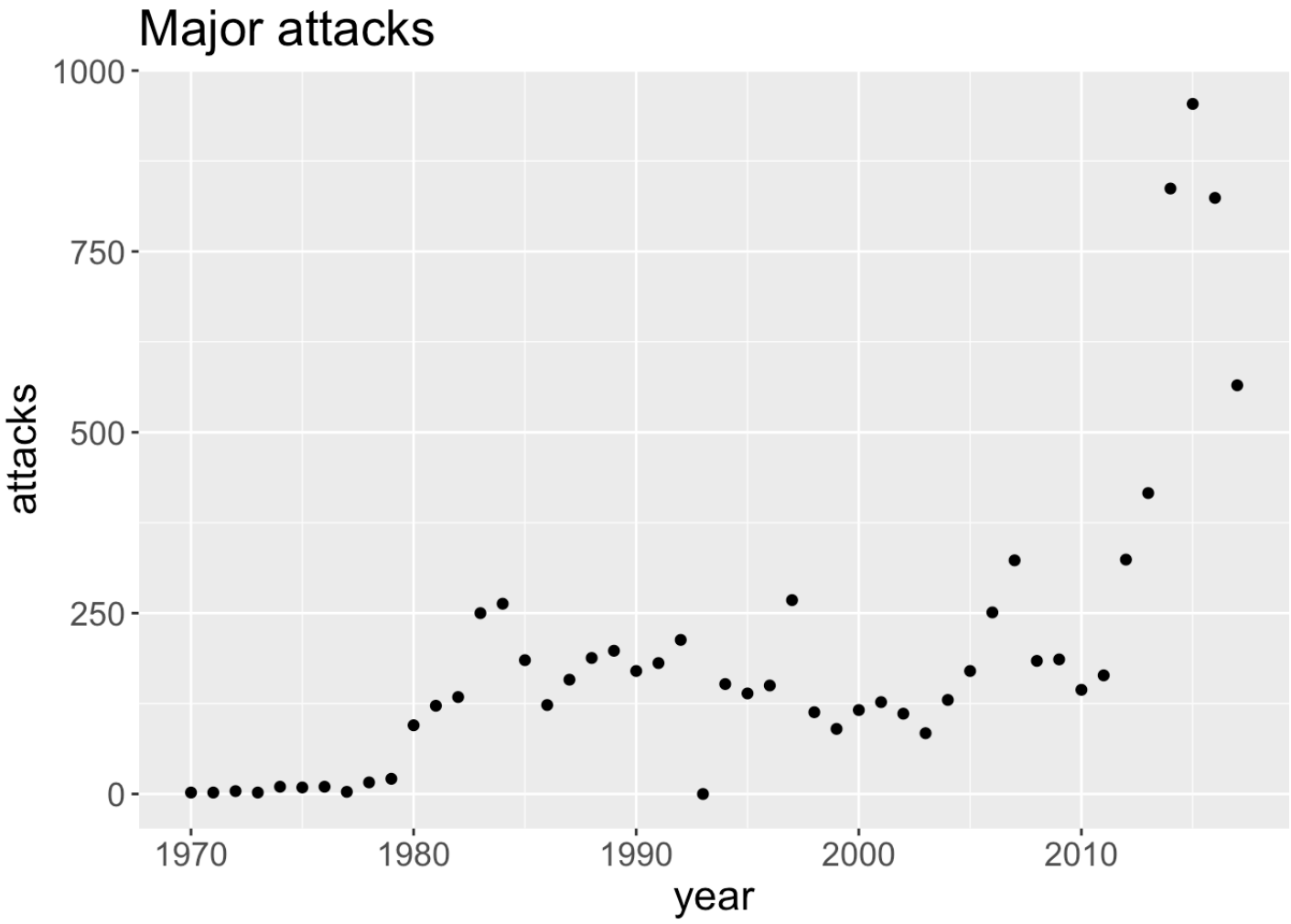
```
ggplot(df1_num, aes(x=year, y=attacks)) + geom_point()+ ggtitle ("Minor attacks") +
        theme(text=element_text(size=16))
```




```
ggplot(df2_num, aes(x=year, y=attacks)) + geom_point() + ggtitle ("Mid-sized attacks"
) + theme(text=element_text(size=16))
```



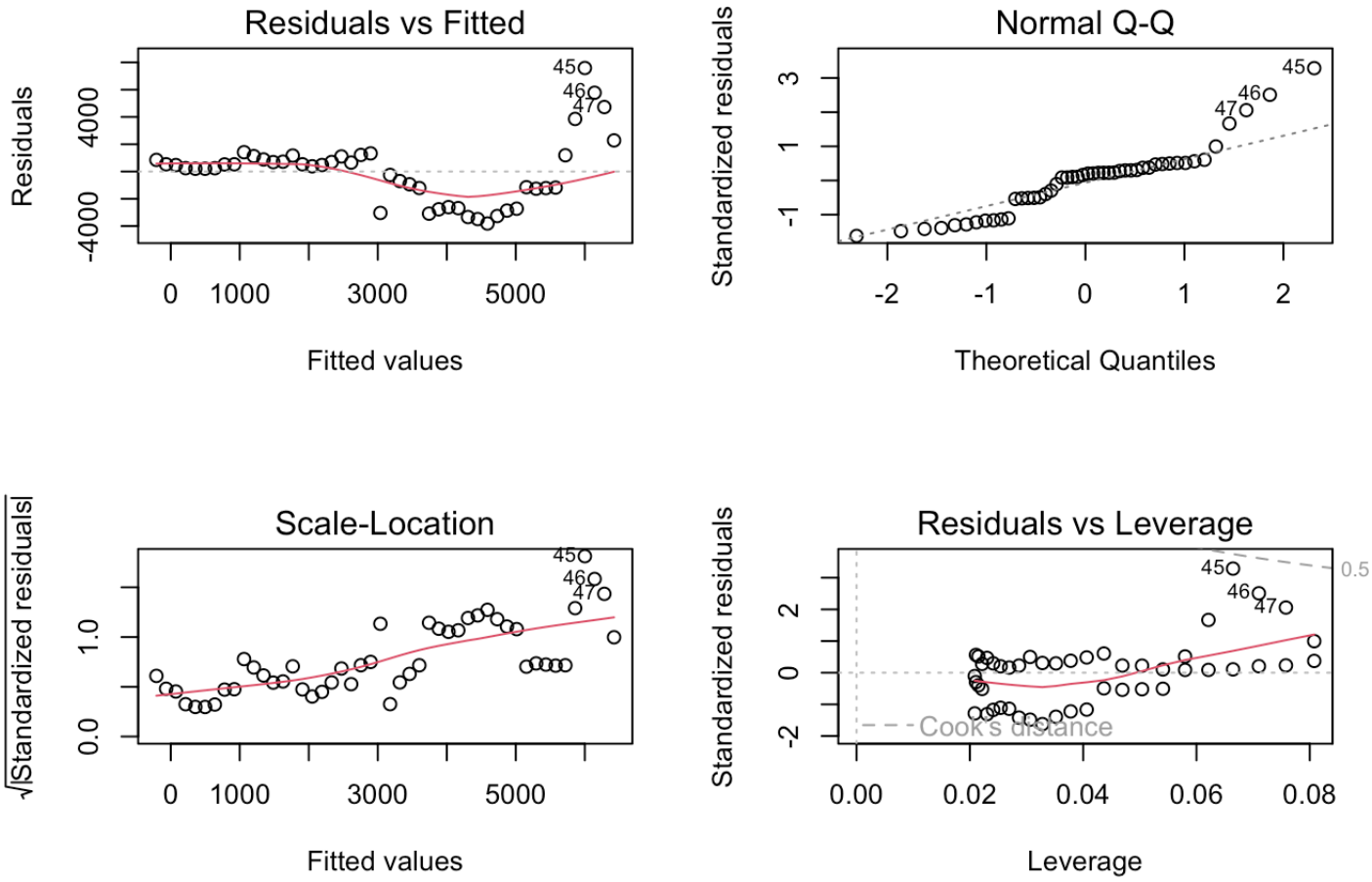

```
ggplot(df3_num, aes(x=year, y=attacks)) + geom_point() + ggtitle ("Major attacks") +
theme(text=element_text(size=16))
```




```
m1 <- lm(attacks ~ year, data=df1_num)
summary(m1)
```

```
##
## Call:
## lm(formula = attacks ~ year, data = df1_num)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3812.6 -1227.2   426.5   938.1  7588.4
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -278182.47   49616.07  -5.607 1.12e-06 ***
## year            141.10      24.89   5.669 9.03e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2389 on 46 degrees of freedom
## Multiple R-squared:  0.4113, Adjusted R-squared:  0.3985
## F-statistic: 32.14 on 1 and 46 DF,  p-value: 9.035e-07
```

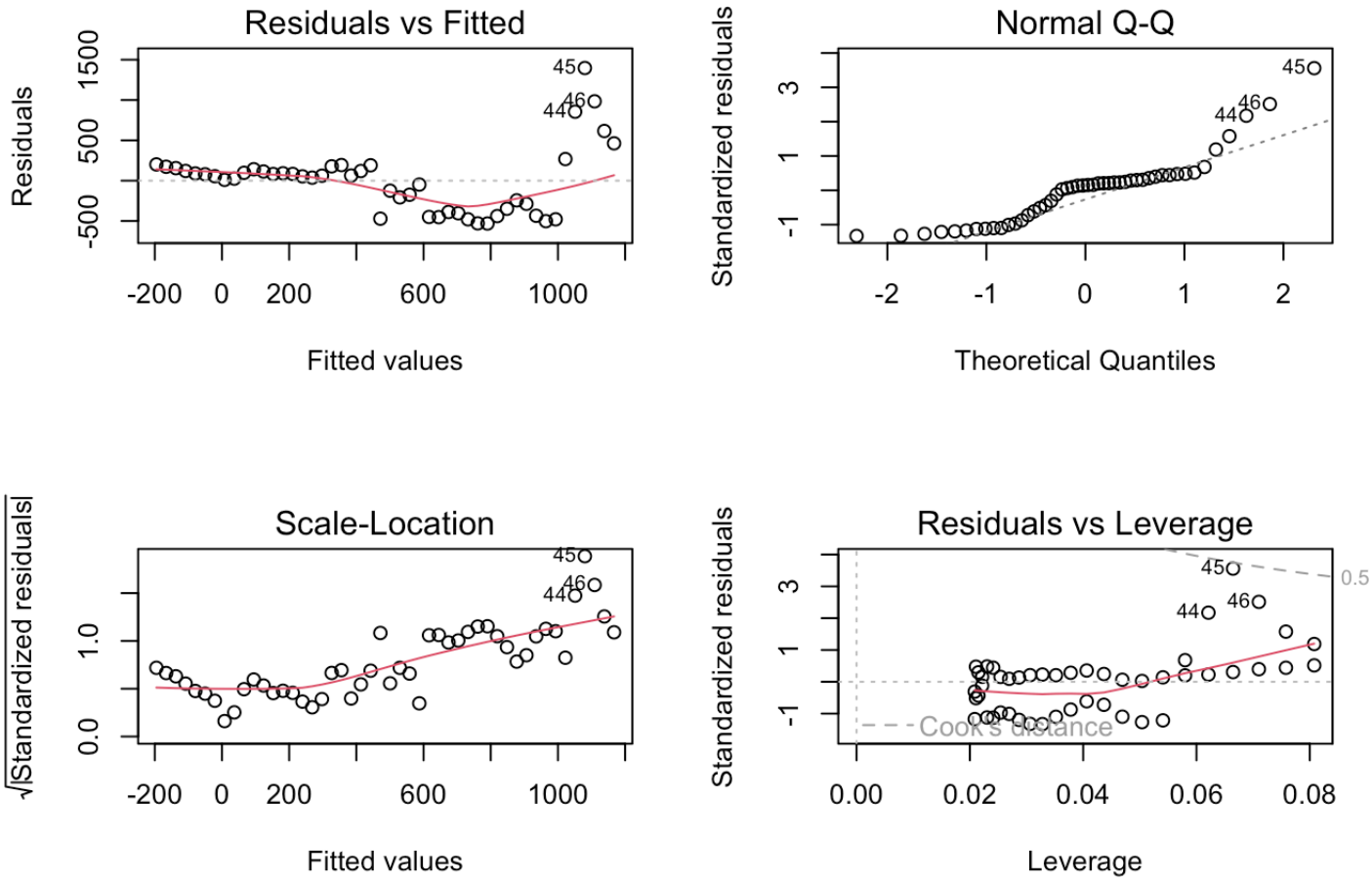
```
par(mfrow= c(2,2))
plot(m1)
```




```
m2 <- lm(attacks ~ year, data=df2_num)
summary(m2)
```

```
##
## Call:
## lm(formula = attacks ~ year, data = df2_num)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -531.28 -358.27   58.38  144.83 1396.87
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -57296.024   8438.849  -6.790 1.89e-08 ***
## year            28.985      4.233   6.847 1.54e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 406.3 on 46 degrees of freedom
## Multiple R-squared:  0.5048, Adjusted R-squared:  0.494
## F-statistic: 46.89 on 1 and 46 DF,  p-value: 1.545e-08
```

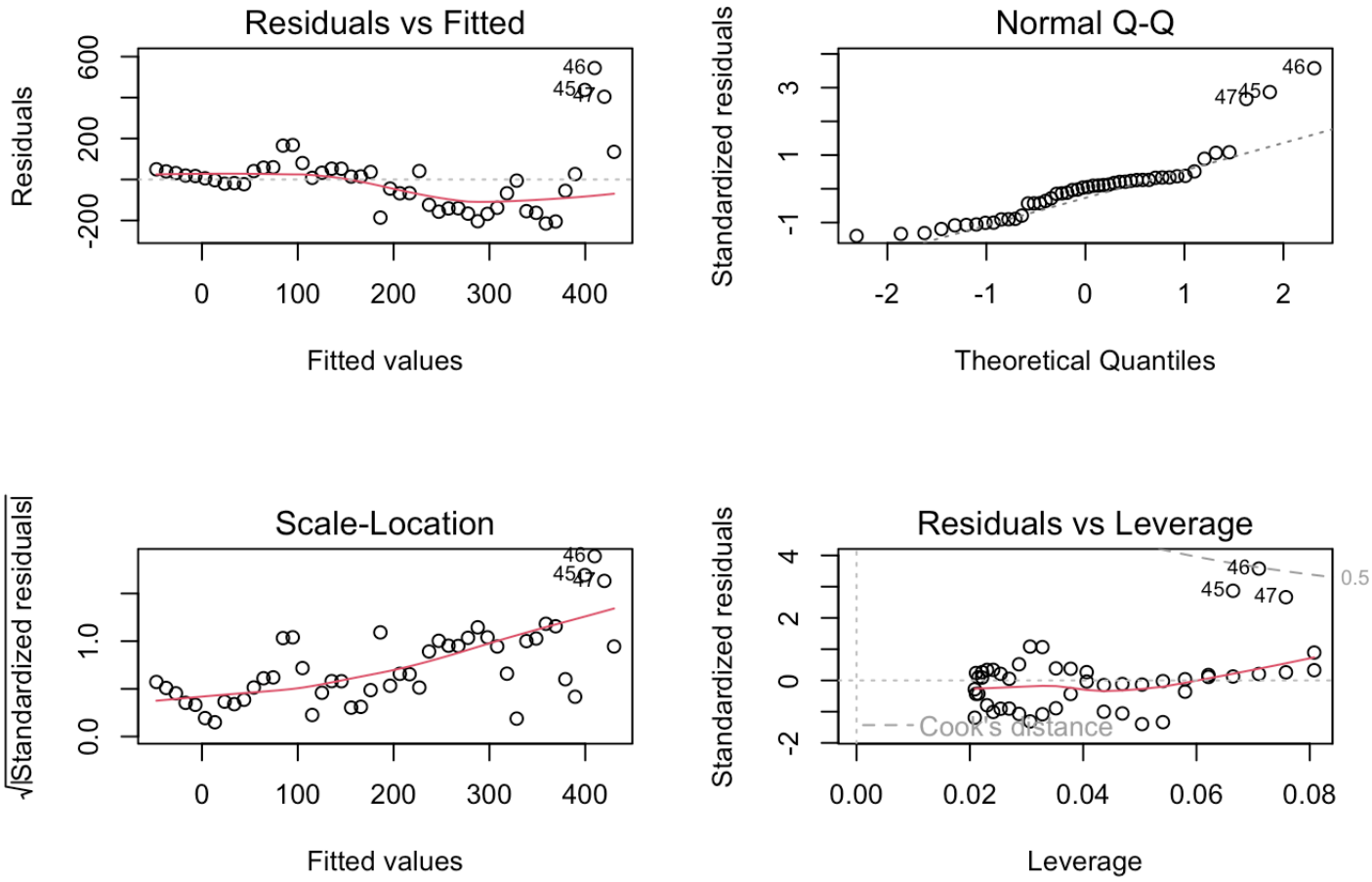
```
par(mfrow= c(2,2))
plot(m2)
```



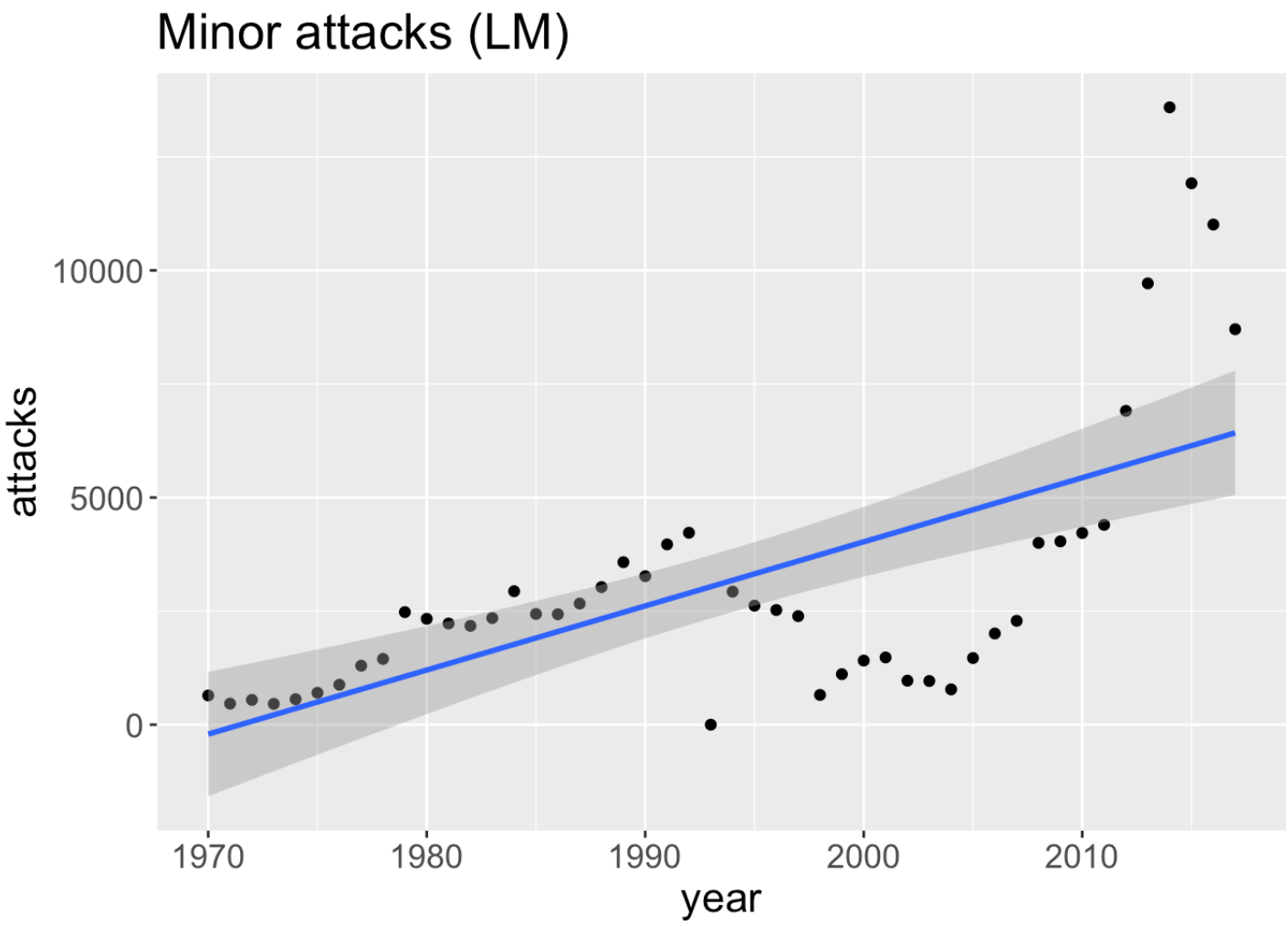

```
m3 <- lm(attacks ~ year, data=df3_num)
summary(m3)
```

```
##
## Call:
## lm(formula = attacks ~ year, data = df3_num)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -214.95 -127.53    6.84   43.26  544.24
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -20066.926   3278.105  -6.122 1.90e-07 ***
## year            10.162      1.644   6.180 1.55e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 157.8 on 46 degrees of freedom
## Multiple R-squared:  0.4536, Adjusted R-squared:  0.4418
## F-statistic: 38.19 on 1 and 46 DF,  p-value: 1.555e-07
```

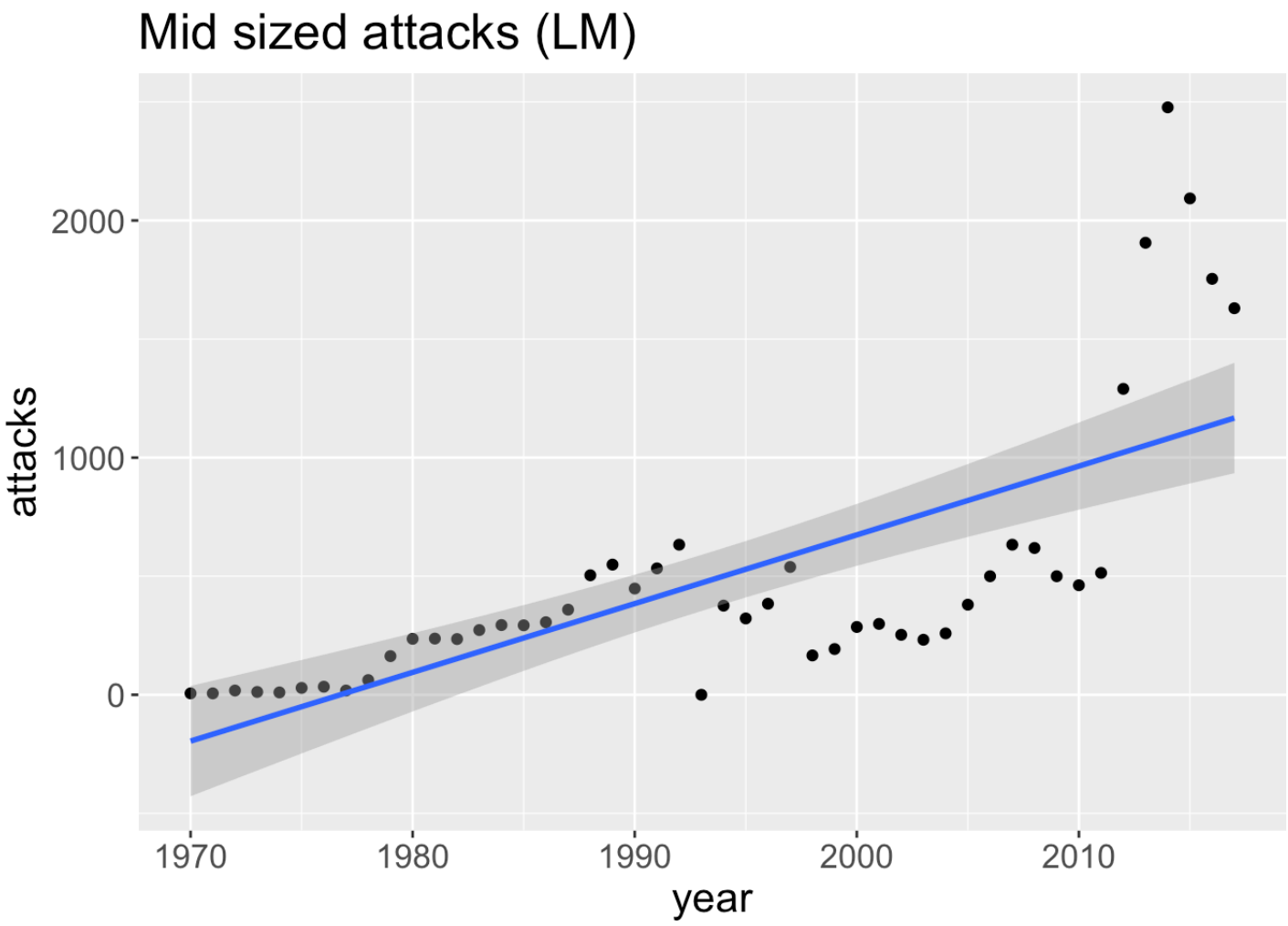
```
par(mfrow= c(2,2))
plot(m3)
```



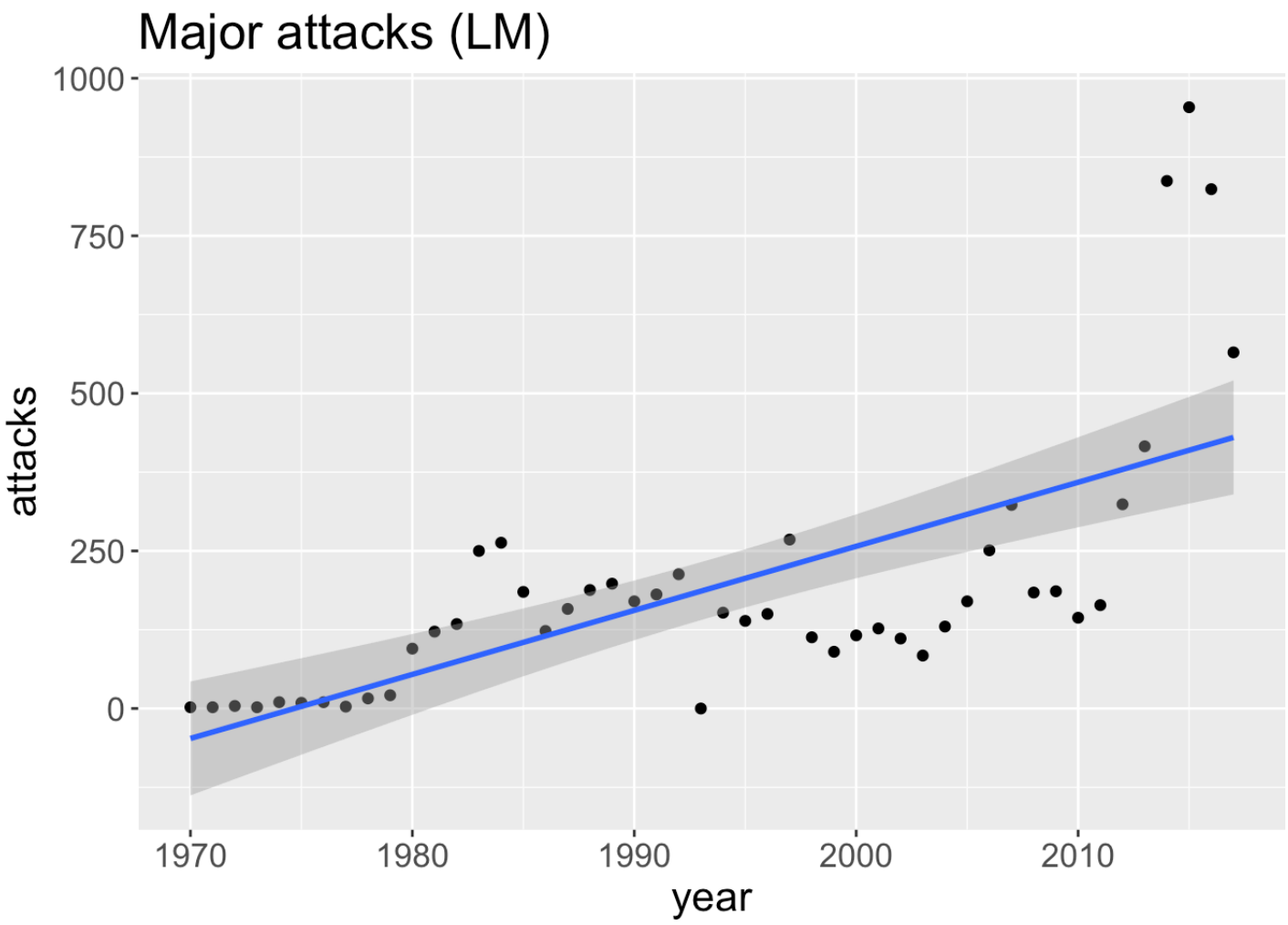

```
ggplot(df1_num, aes(x=year, y=attacks)) + geom_point() + ggtitle ("Minor attacks (LM)
") + theme(text=element_text(size=16)) +
stat_smooth(method = "lm",
              formula = y ~ x,
              geom = "smooth")
```



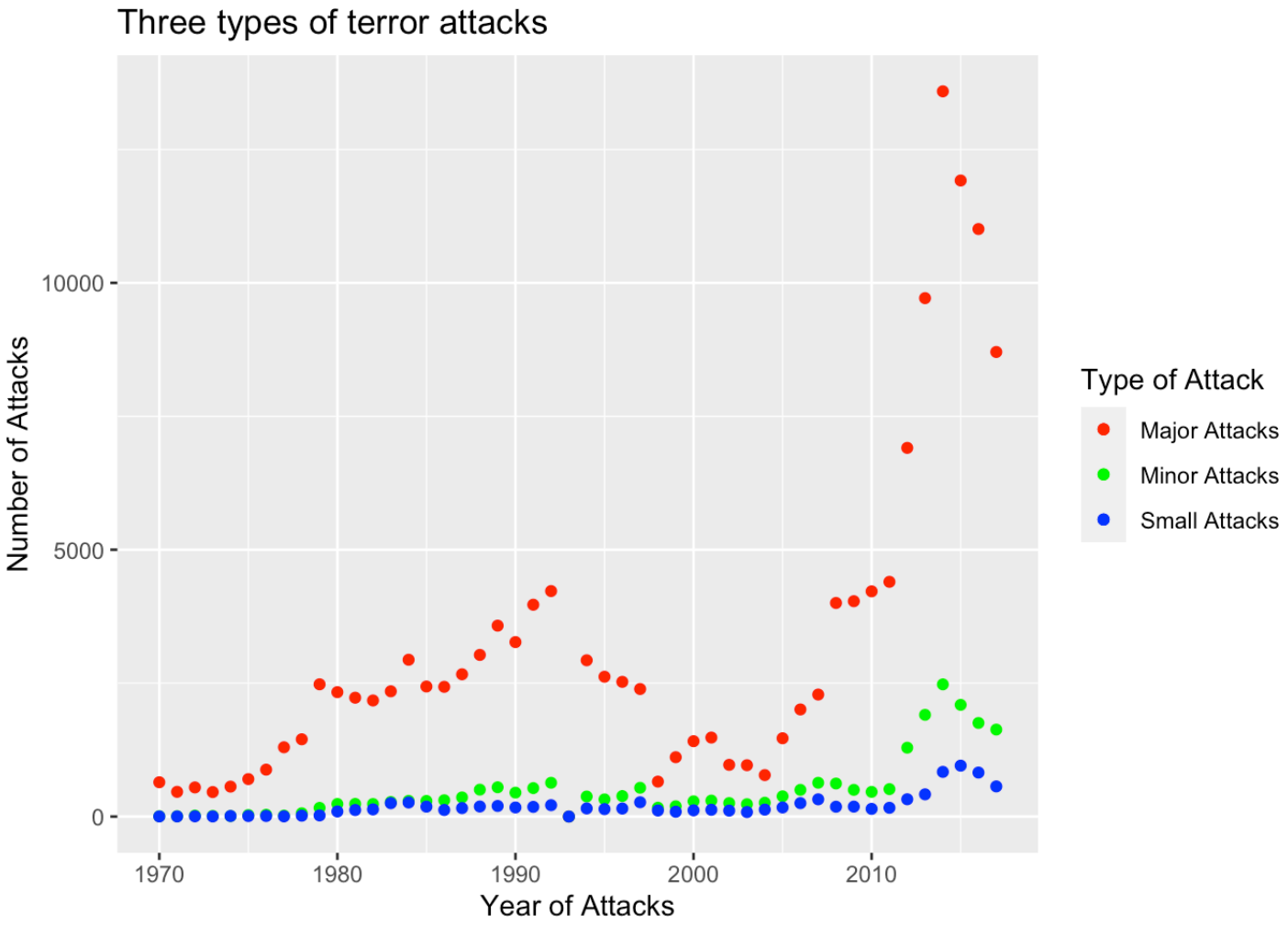

```
ggplot(df2_num, aes(x=year, y=attacks)) + geom_point() + ggtitle ("Mid sized attacks
(LM)") + theme(text=element_text(size=16)) +
stat_smooth(method = "lm",
              formula = y ~ x,
              geom = "smooth")
```




```
ggplot(df3_num, aes(x=year, y=attacks)) + geom_point() + ggtitle ("Major attacks (LM)
") + theme(text=element_text(size=16)) +
stat_smooth(method = "lm",
              formula = y ~ x,
              geom = "smooth")
```




```
colors <- c("Major Attacks"="red","Small Attacks"="blue","Minor Attacks"="green")
ggplot()+
  geom_point(data=df1_num, aes(x=year,y=attacks,color="Major Attacks"), show.legend =
TRUE)+
  geom_point(data=df2_num, aes(x=year,y=attacks,color="Minor Attacks"),show.legend =
TRUE)+
  geom_point(data=df3_num, aes(x=year,y=attacks,color="Small Attacks"),show.legend =
TRUE)+
  ggtitle('Three types of terror attacks') +
  labs(color = "Type of Attack")+
  scale_color_manual(values = colors)+
  scale_x_continuous("Year of Attacks")+
  scale_y_continuous("Number of Attacks")
```



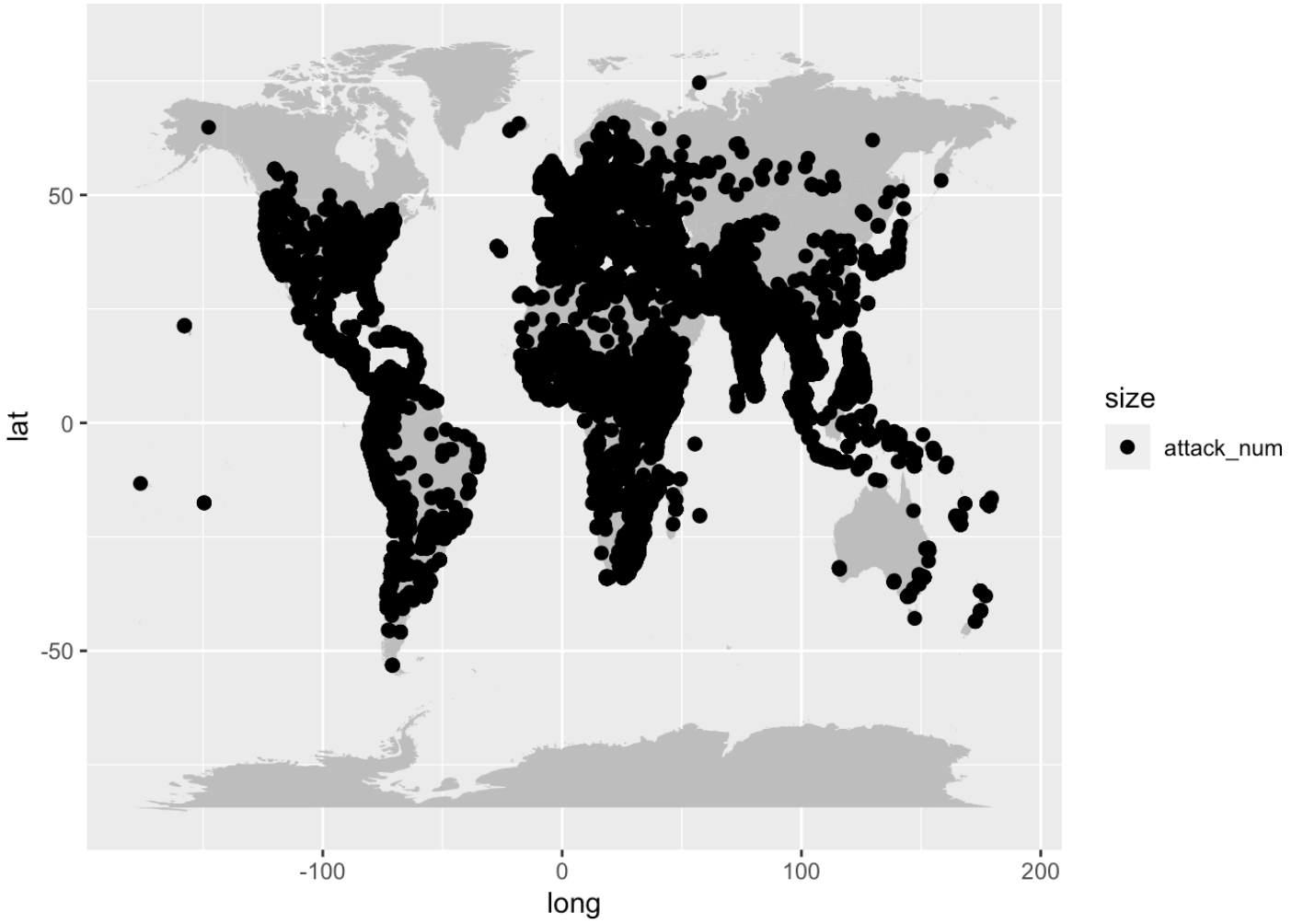

```
library(tidyverse)
library(dplyr)
library(ggplot2)
```

```
world_data <- raw_data
world_data1 <- world_data[!is.na(world_data$latitude),]
world_data2 <- world_data1[!is.na(world_data1$longitude),]

world_data2$attack_var <- world_data2$nkill

world_data2 <- world_data2 %>% mutate(attack_var = replace(attack_var, world_data2$nk
ill>=3 & world_data2$nkill<10,2))
world_data2 <- world_data2 %>% mutate(attack_var = replace(attack_var, world_data2$nk
ill>=10,3))
world_data2 <- world_data2 %>% mutate(attack_var = replace(attack_var, world_data2$nk
ill<3,1))

world_data2[is.na(world_data2)] <- 0

wd1 <- world_data2[world_data2$attack_var == "1", ]
wd2 <- world_data2[world_data2$attack_var == "2", ]
wd3 <- world_data2[world_data2$attack_var == "3", ]
```

```
world_coordinates <- map_data("world")
options(repr.plot.width = 50, repr.plot.height =50)

ggplot() + geom_map(
    data = world_coordinates, map = world_coordinates,
    aes(long, lat, map_id = region), fill="grey"
  ) +
geom_point(
    data = wd1,
    aes(x=longitude, y=latitude, size='attack_num')
  )
```

```
## Warning in geom_map(data = world_coordinates, map = world_coordinates,
## aes(long, : Ignoring unknown aesthetics: x and y
```

```
## Warning: Using size for a discrete variable is not advised.
```




```
library(ggmap)
```

```
## ℹ Google's Terms of Service: <https://mapsplatform.google.com>
## ℹ Please cite ggmap if you use it! Use `citation("ggmap")` for details.
```

```
world <- map_data("world")
```

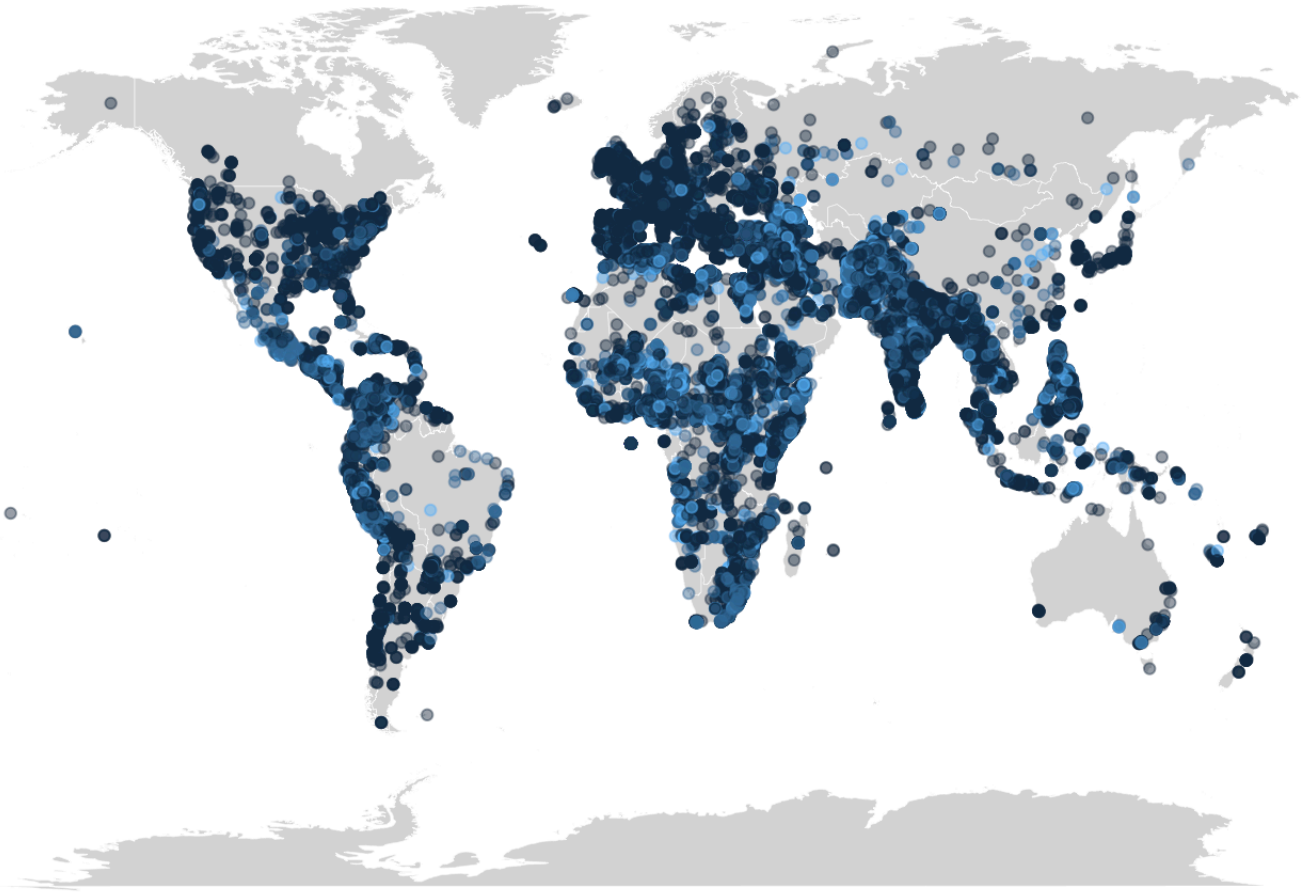
Major attacks

```
ggplot() +
  geom_map(
    data = world, map = world,
    aes(long, lat, map_id = region),
    color = "white", fill = "lightgray", size = 0.1
  ) +
  geom_point(
    data = df1,
    aes(longitude, latitude,
        color = nkill),
    alpha = 0.5
  ) +
  labs(x = NULL, y = NULL, color = NULL)+
  theme_void() +
  theme(legend.position = "none")+
  labs(title="Major Terror Attack Locations")
```

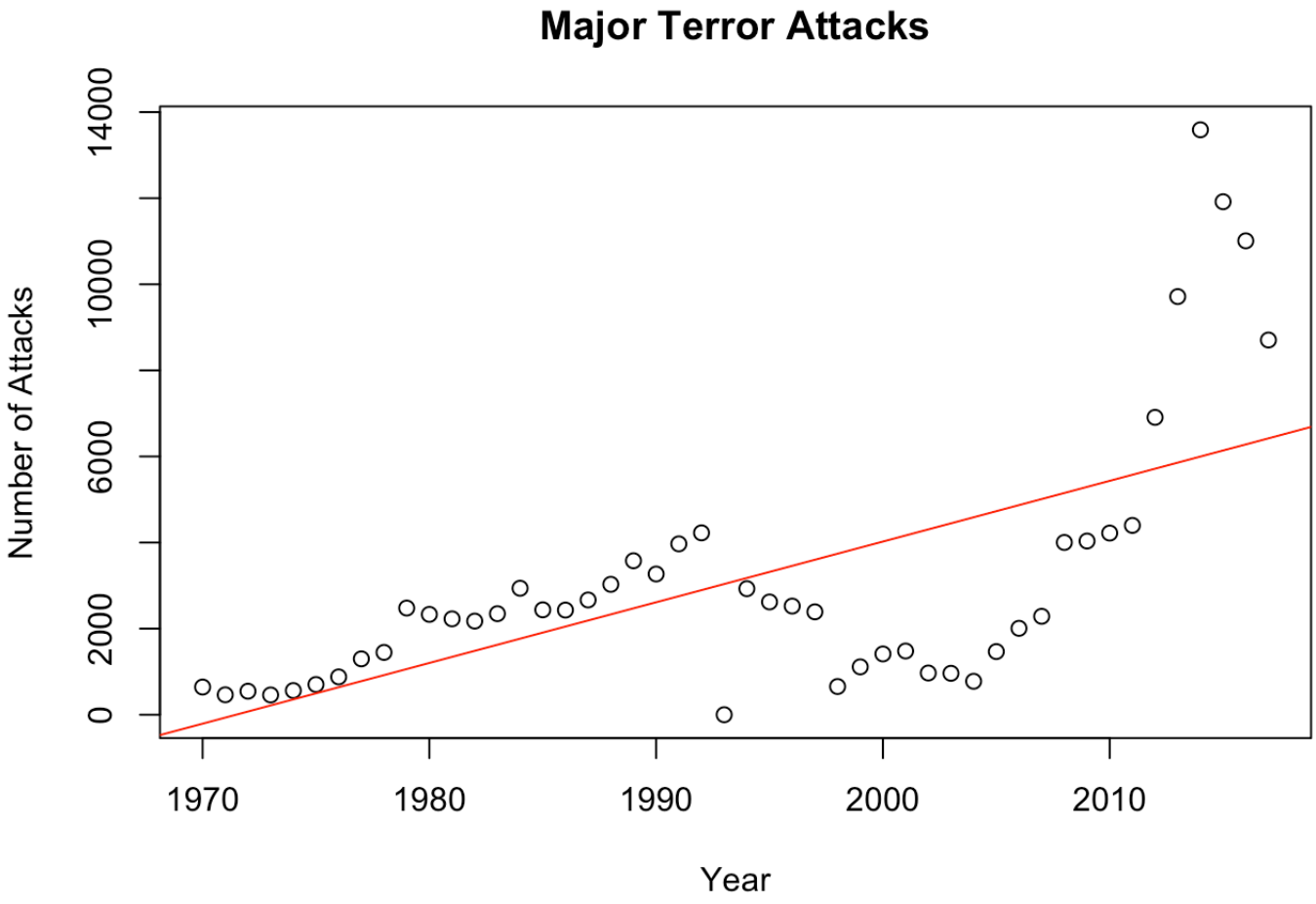
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## Warning in geom_map(data = world, map = world, aes(long, lat, map_id =
## region), : Ignoring unknown aesthetics: x and y
```

Major Terror Attack Locations



```
plot(x=df1_num$year,y=df1_num$attacks,main = "Major Terror Attacks",
     xlab = "Year",
     ylab = "Number of Attacks")
abline(m1,col = "red")
```




```
library(car)
```

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

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```
# Calculating Residuals
residuals <- m1$residuals

# Normality test
ks.test(residuals, rnorm(100,0,1))    # Since p value is very small, we can reject th
e null hypothesis
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  residuals and rnorm(100, 0, 1)
## D = 0.60417, p-value = 1.275e-11
## alternative hypothesis: two-sided
```

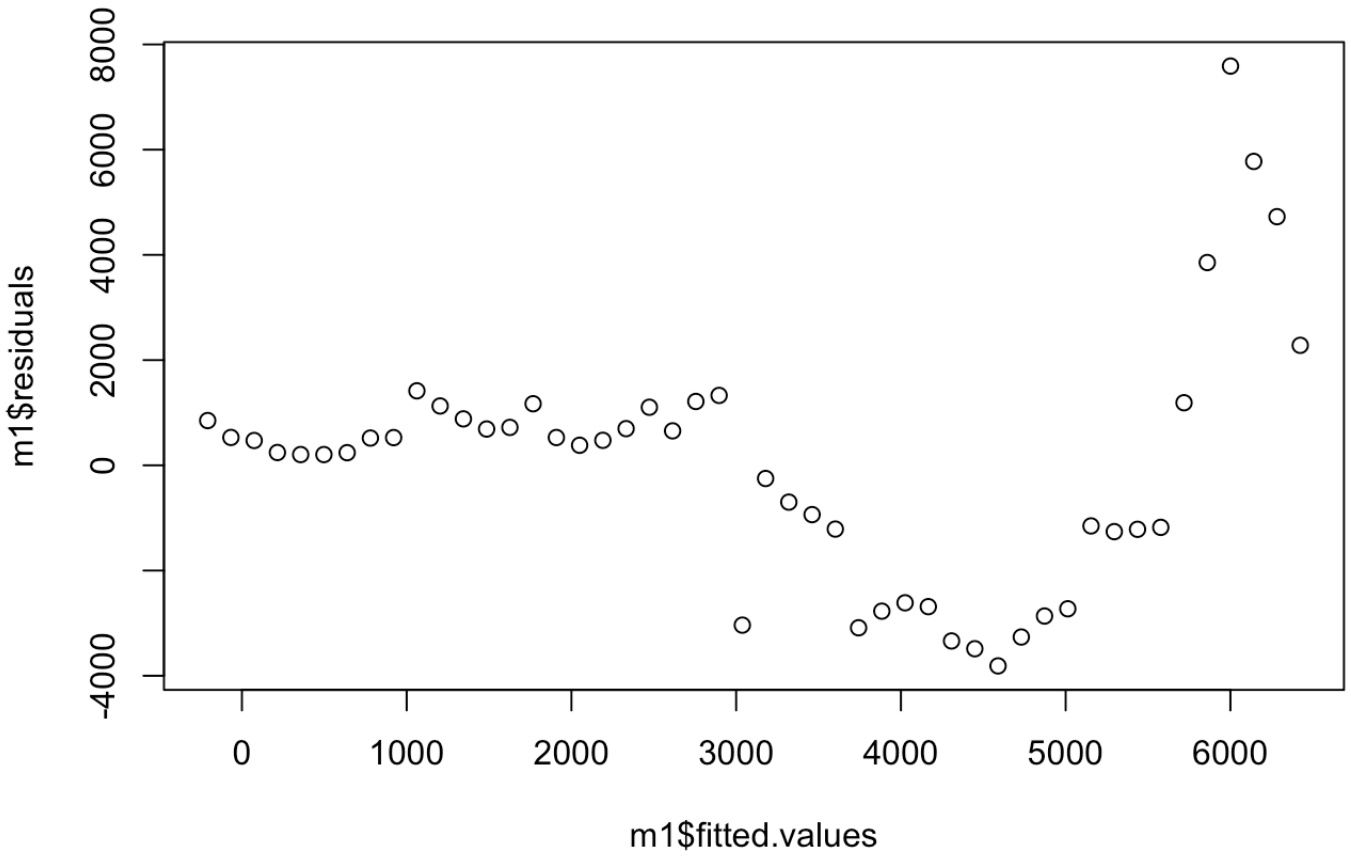
```
# And say that residuals do not follow a normal distritbution
shapiro.test(residuals)               # Small value of Shapiro-Wilk test also leads u
p to the same result
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals
## W = 0.90946, p-value = 0.001288
```

```
# Constant Variance Assumption/Cook Weinberg Test
ncvTest(m1)                          # Small value indicates that assumption is violat
ed
```

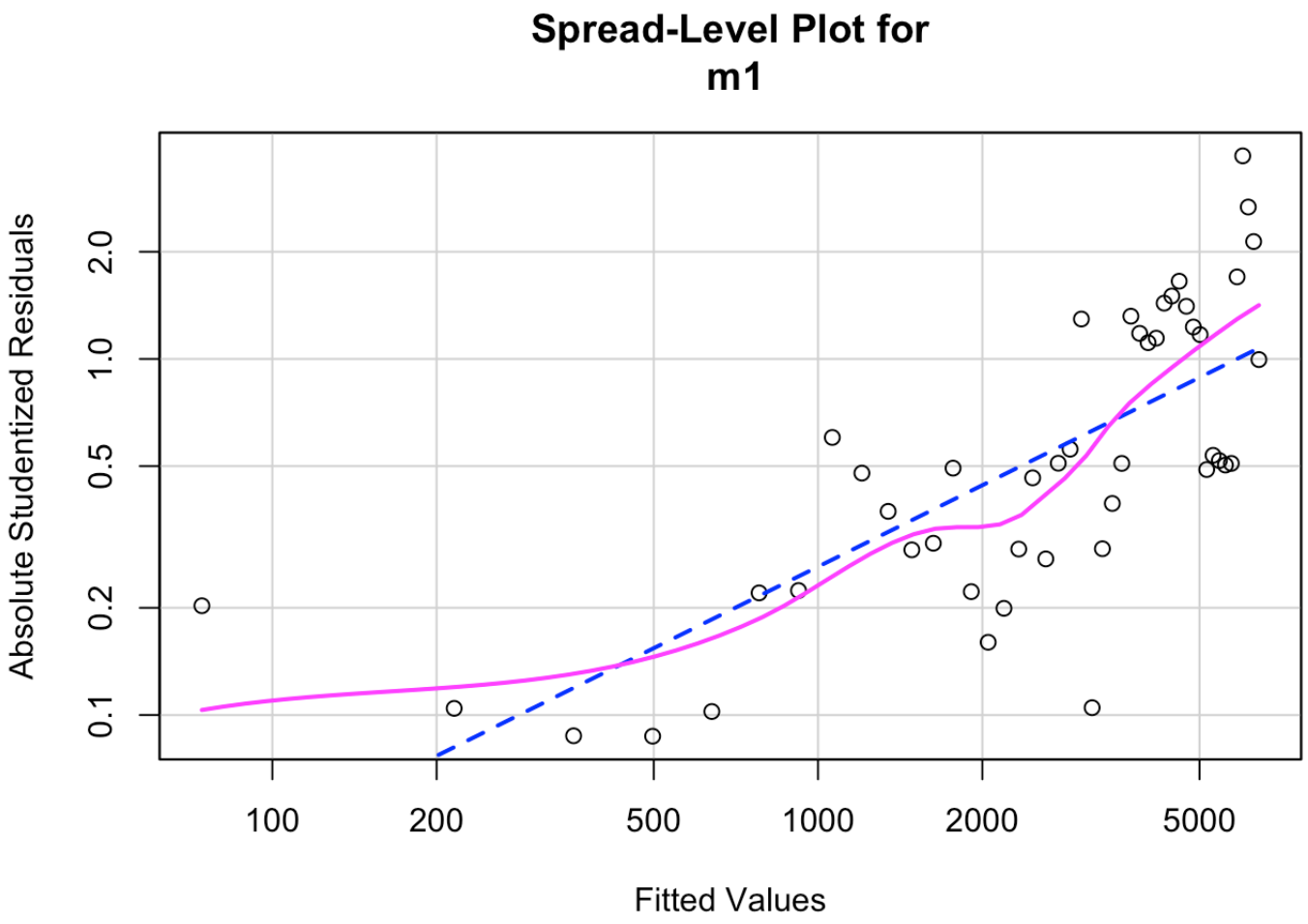
```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 24.3818, Df = 1, p = 7.9012e-07
```

```
plot(m1$fitted.values, m1$residuals)
```




```
# Spread Level Plot
myspread <- spreadLevelPlot(m1)
```

```
## Warning in spreadLevelPlot.lm(m1):
## 2 negative fitted values removed
```




```
myspread
```

```
##
## Suggested power transformation:  0.2399187
```

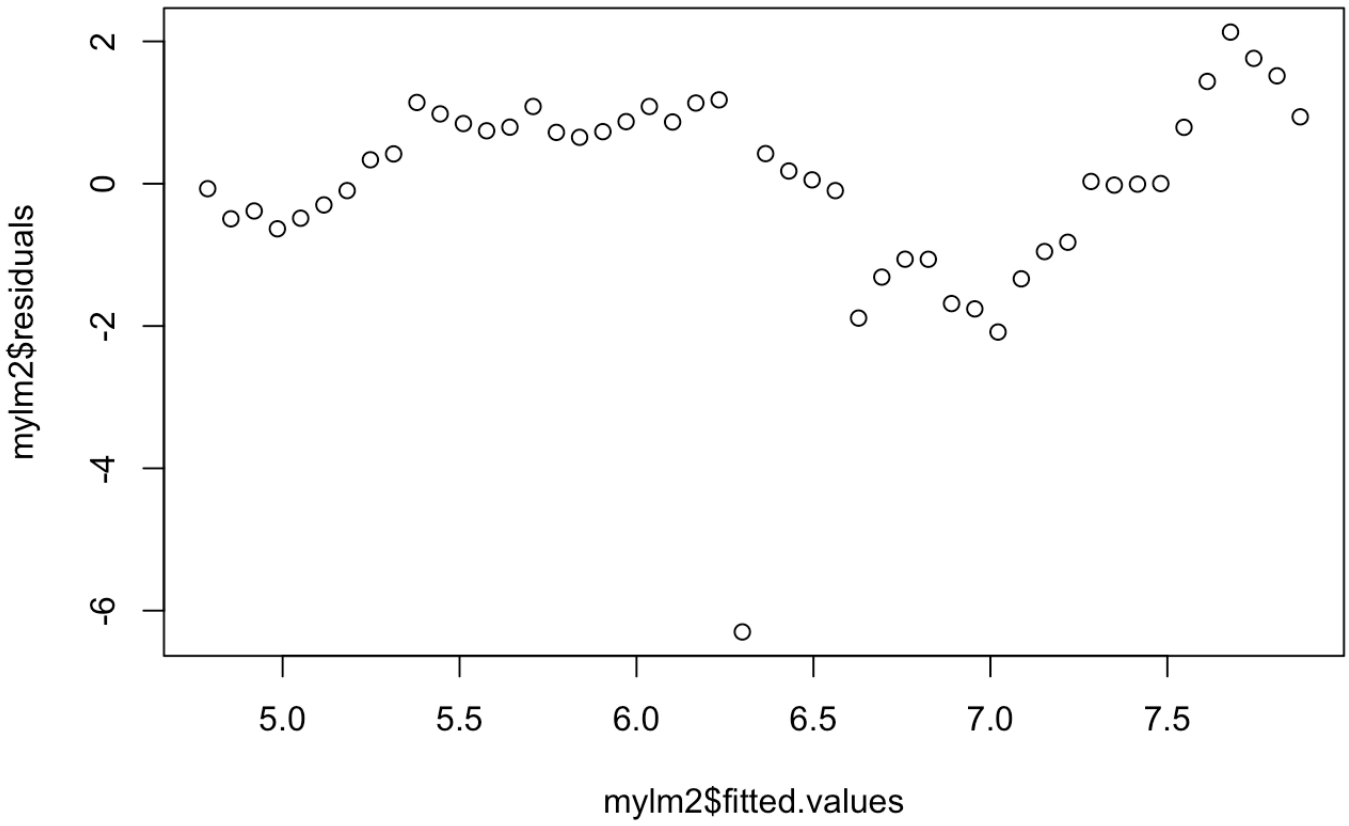
```
y <- df1_num$attacks
x <- df1_num$year

# Spread Level transformation and New Model
z<-y^(myspread$PowerTransformation)
mylm2<-lm(z ~ x)
summary(mylm2)
```

```
##
## Call:
## lm(formula = z ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.2990 -0.5286  0.1160  0.8684  2.1307
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -124.61586   28.76360  -4.332 7.93e-05 ***
## x              0.06569    0.01443   4.553 3.88e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.385 on 46 degrees of freedom
## Multiple R-squared:  0.3106, Adjusted R-squared:  0.2956
## F-statistic: 20.73 on 1 and 46 DF,  p-value: 3.876e-05
```

```
# Cook Weinberg Test Again
ncvTest(mylm2)                      # a high value of p indicates constant variance
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 1.692439, Df = 1, p = 0.19328
```

```
plot(mylm2$fitted.values, mylm2$residuals)
```




```
# Normality Test
residuals2 <- mylm2$residuals
n <- 47
sd1 <- sd(mylm2$residuals)

ks.test(rnorm(n,0,sd1),mylm2$residuals)   # Large value of p indicates that residuals
are normal
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  rnorm(n, 0, sd1) and mylm2$residuals
## D = 0.15691, p-value = 0.5339
## alternative hypothesis: two-sided
```

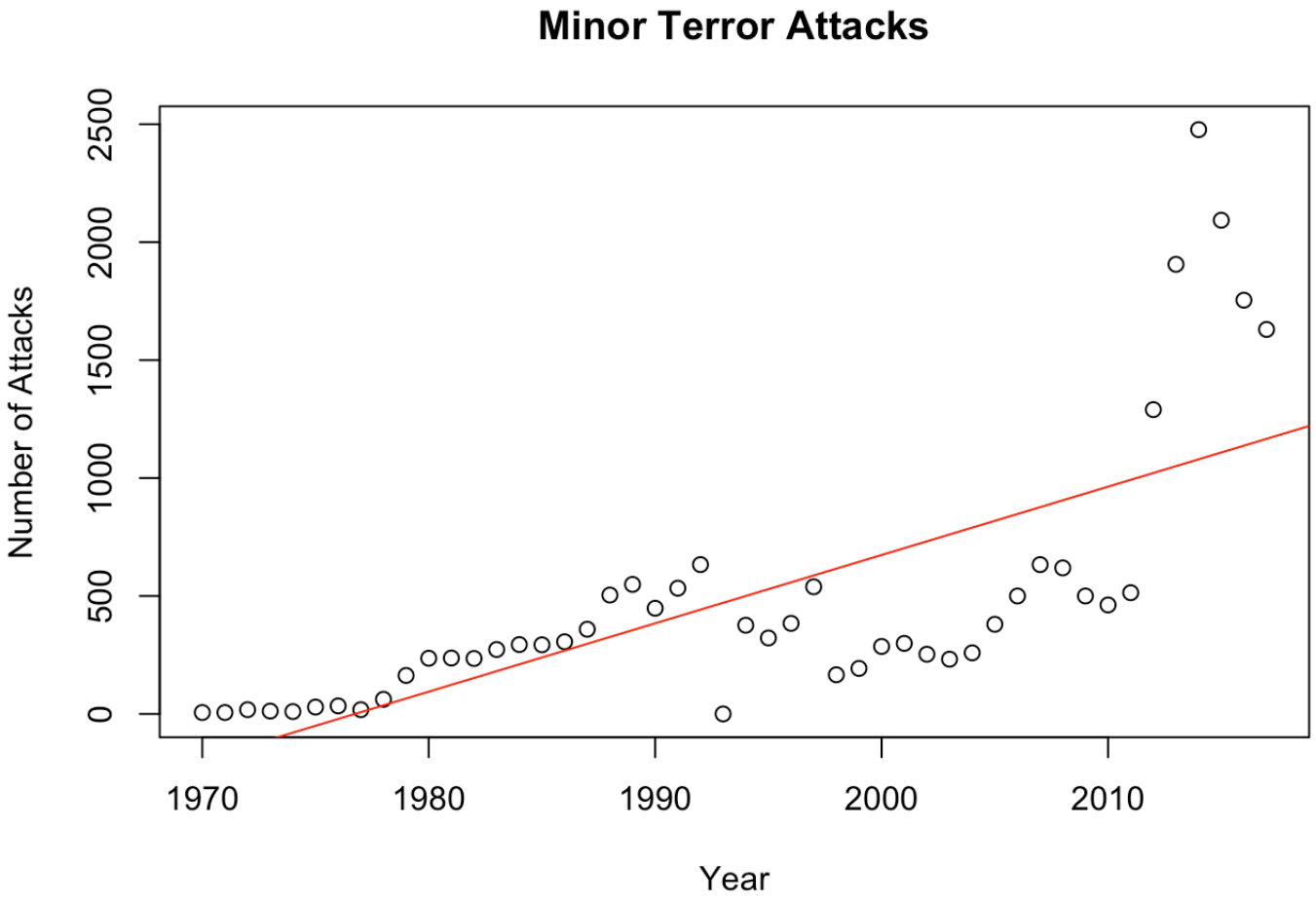
Minor attacks

```
ggplot() +
  geom_map(
    data = world, map = world,
    aes(long, lat, map_id = region),
    color = "white", fill = "lightgray", size = 0.1
  ) +
  geom_point(
    data = df2,
    aes(longitude, latitude,
        color = nkill),
    alpha = 0.5
  ) +
  labs(x = NULL, y = NULL, color = NULL)+
  theme_void() +
  theme(legend.position = "none")+
  labs(title="Minor Terror Attack Locations")
```

```
## Warning in geom_map(data = world, map = world, aes(long, lat, map_id =
## region), : Ignoring unknown aesthetics: x and y
```

Minor Terror Attack Locations

```
plot(x=df2_num$year,y=df2_num$attacks,main = "Minor Terror Attacks",
     xlab = "Year",
     ylab = "Number of Attacks")
abline(m2,col = "red")
```




```
# Calculating Residuals
residuals <- m2$residuals

# Normality test
ks.test(residuals, rnorm(100,0,1))    # Since p value is very small, we can reject th
e null hypothesis
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  residuals and rnorm(100, 0, 1)
## D = 0.60417, p-value = 1.275e-11
## alternative hypothesis: two-sided
```

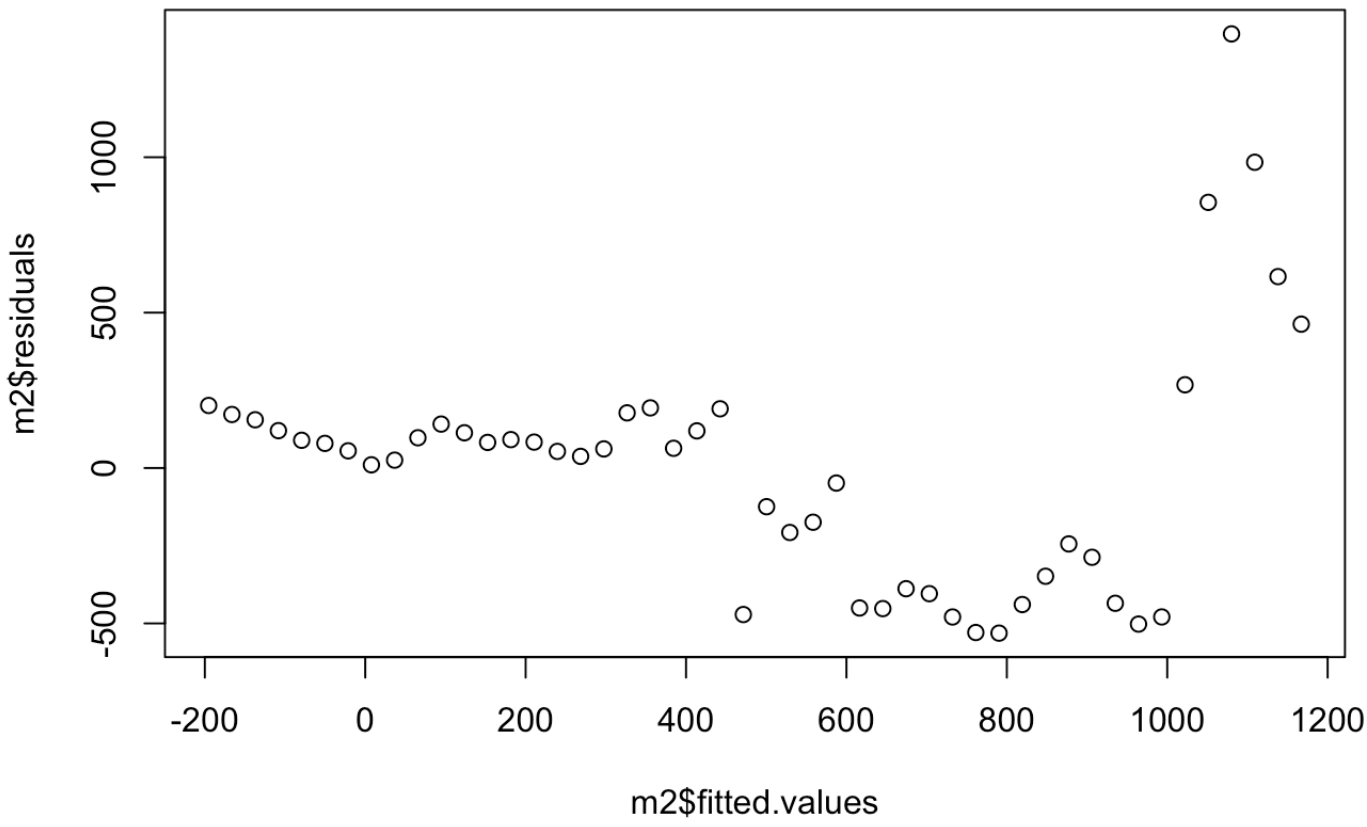
```
# And say that residuals do not follow a normal distritbution
shapiro.test(residuals)               # Small value of Shapiro-Wilk test also leads u
p to the same result
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals
## W = 0.8801, p-value = 0.0001522
```

```
# Constant Variance Assumption/Cook Weinberg Test
ncvTest(m2)                          # Small value indicates that assumption is violat
ed
```

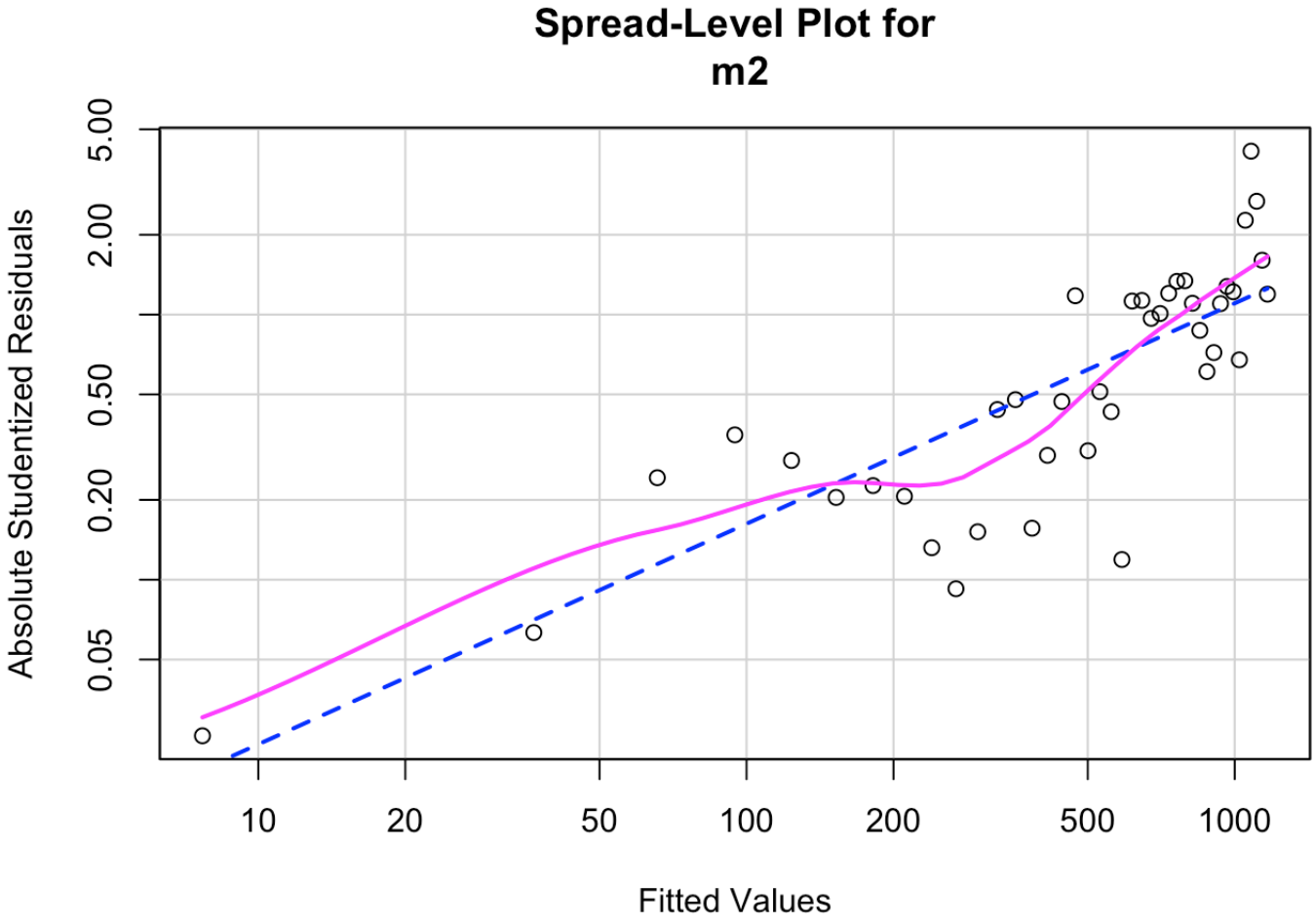
```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 28.48745, Df = 1, p = 9.4308e-08
```

```
plot(m2$fitted.values, m2$residuals)
```



```
# Spread Level Plot
myspread <- spreadLevelPlot(m2)
```

```
## Warning in spreadLevelPlot.lm(m2):
## 7 negative fitted values removed
```




```
myspread
```

```
##
## Suggested power transformation:  0.1683115
```

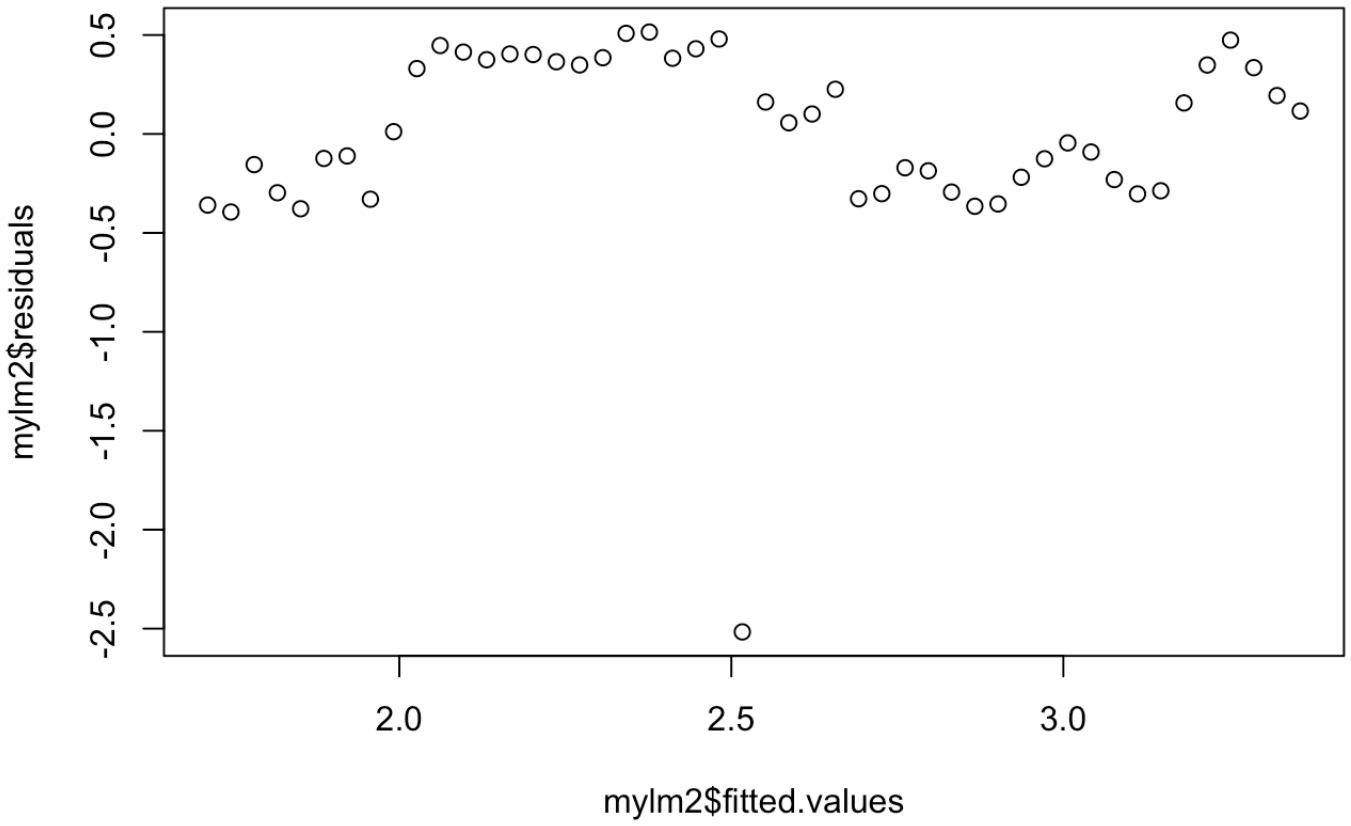
```
y <- df2_num$attacks
x <- df2_num$year

# Spread Level transformation and New Model
z<-y^(myspread$PowerTransformation)
mylm2<-lm(z ~ x)
summary(mylm2)
```

```
##
## Call:
## lm(formula = z ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.51658 -0.28888  0.03398  0.36740  0.51480
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -67.25423   10.14640  -6.628 3.29e-08 ***
## x             0.03501    0.00509   6.878 1.39e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4885 on 46 degrees of freedom
## Multiple R-squared:  0.507,  Adjusted R-squared:  0.4963
## F-statistic: 47.31 on 1 and 46 DF,  p-value: 1.388e-08
```

```
# Cook Weinberg Test Again
ncvTest(mylm2)                      # a high value of p indicates constant variance
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.2236473, Df = 1, p = 0.63627
```

```
plot(mylm2$fitted.values, mylm2$residuals)
```




```
# Normality Test
residuals2 <- mylm2$residuals
n <- 47
sd1 <- sd(mylm2$residuals)

ks.test(rnorm(n,0,sd1),mylm2$residuals)   # Large value of p indicates that residuals
are normal
```

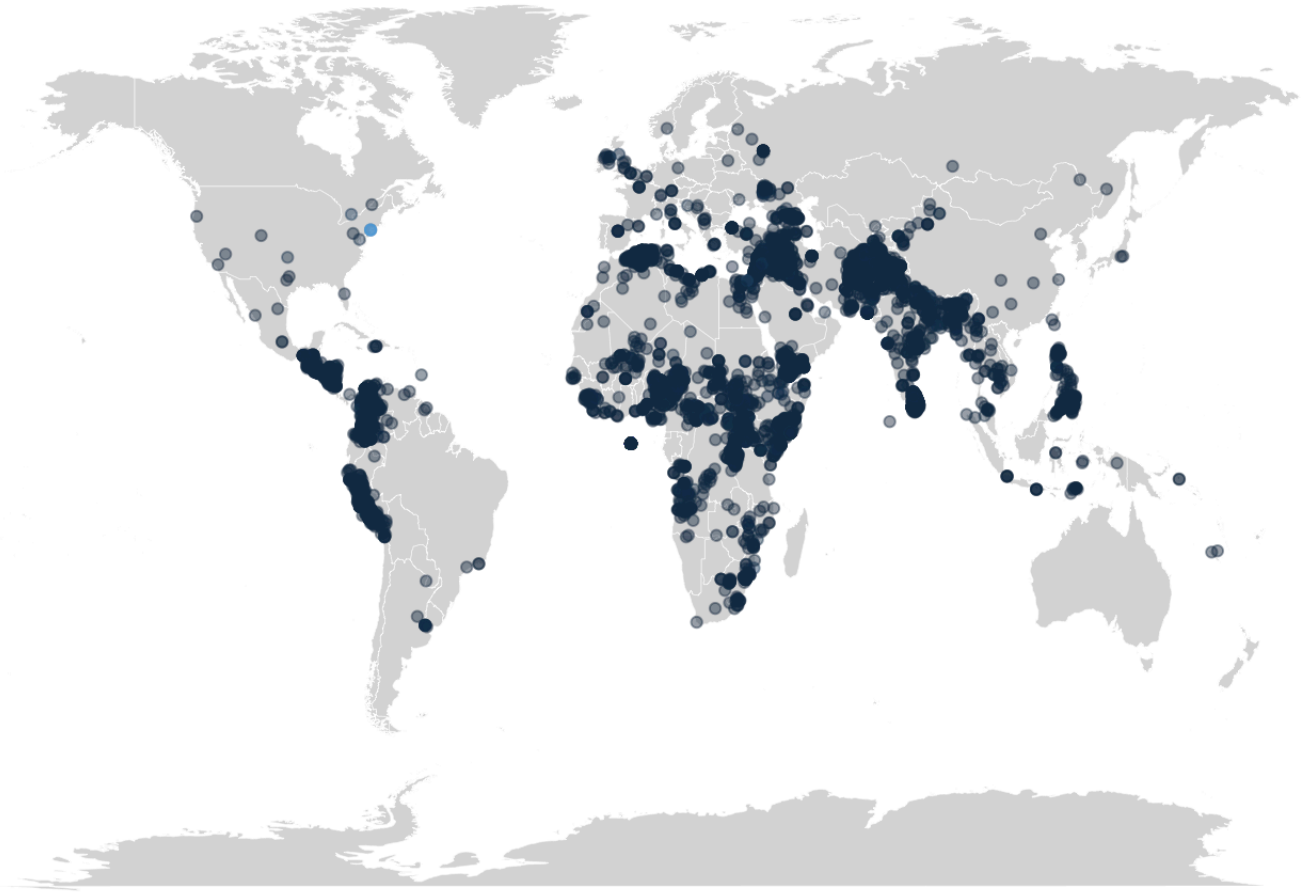
```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  rnorm(n, 0, sd1) and mylm2$residuals
## D = 0.19193, p-value = 0.3017
## alternative hypothesis: two-sided
```

Small attacks

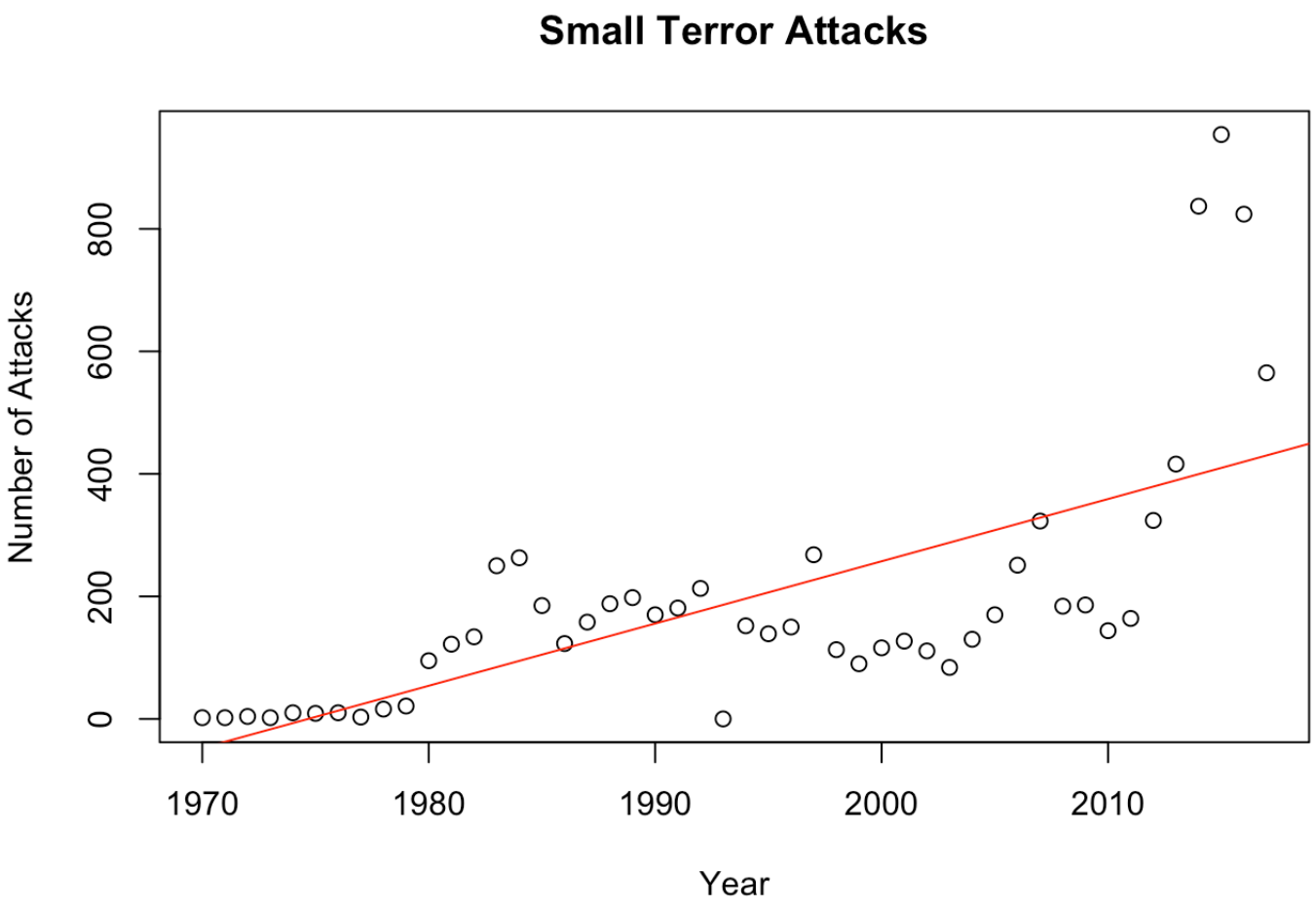
```
ggplot() +
  geom_map(
    data = world, map = world,
    aes(long, lat, map_id = region),
    color = "white", fill = "lightgray", size = 0.1
  ) +
  geom_point(
    data = df3,
    aes(longitude, latitude,
        color = nkill),
    alpha = 0.5
  ) +
  labs(x = NULL, y = NULL, color = NULL)+
  theme_void() +
  theme(legend.position = "none")+
  labs(title="Small Terror Attack Locations")
```

```
## Warning in geom_map(data = world, map = world, aes(long, lat, map_id =
## region), : Ignoring unknown aesthetics: x and y
```

Small Terror Attack Locations



```
plot(x=df3_num$year,y=df3_num$attacks,main = "Small Terror Attacks",
     xlab = "Year",
     ylab = "Number of Attacks")
abline(m3,col = "red")
```




```
# Calculating Residuals
residuals <- m3$residuals

# Normality test
ks.test(residuals, rnorm(100,0,1))    # Since p value is very small, we can reject th
e null hypothesis
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  residuals and rnorm(100, 0, 1)
## D = 0.52083, p-value = 1.384e-08
## alternative hypothesis: two-sided
```

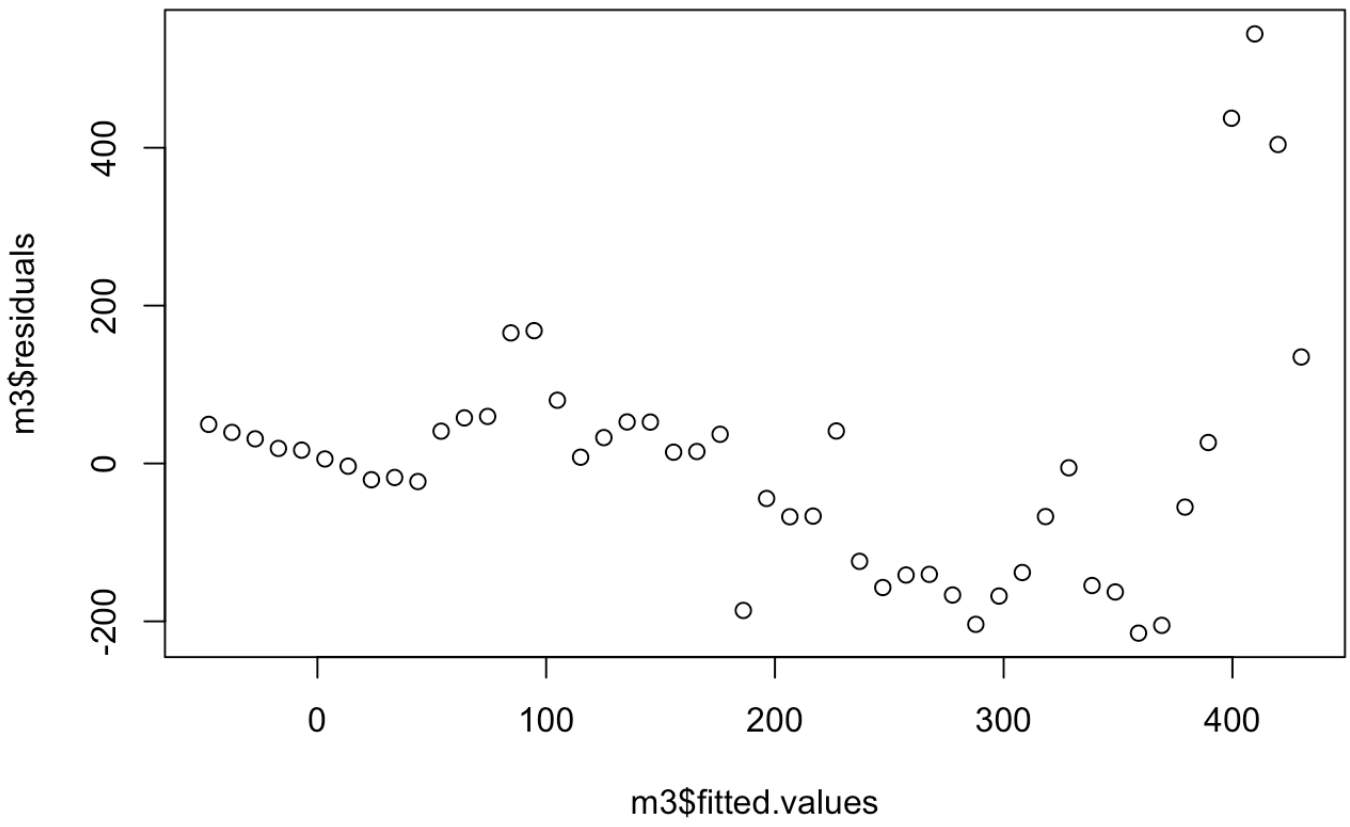
```
# And say that residuals do not follow a normal distritbution
shapiro.test(residuals)               # Small value of Shapiro-Wilk test also leads u
p to the same result
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals
## W = 0.85464, p-value = 2.895e-05
```

```
# Constant Variance Assumption/Cook Weinberg Test
ncvTest(m3)                          # Small value indicates that assumption is violat
ed
```

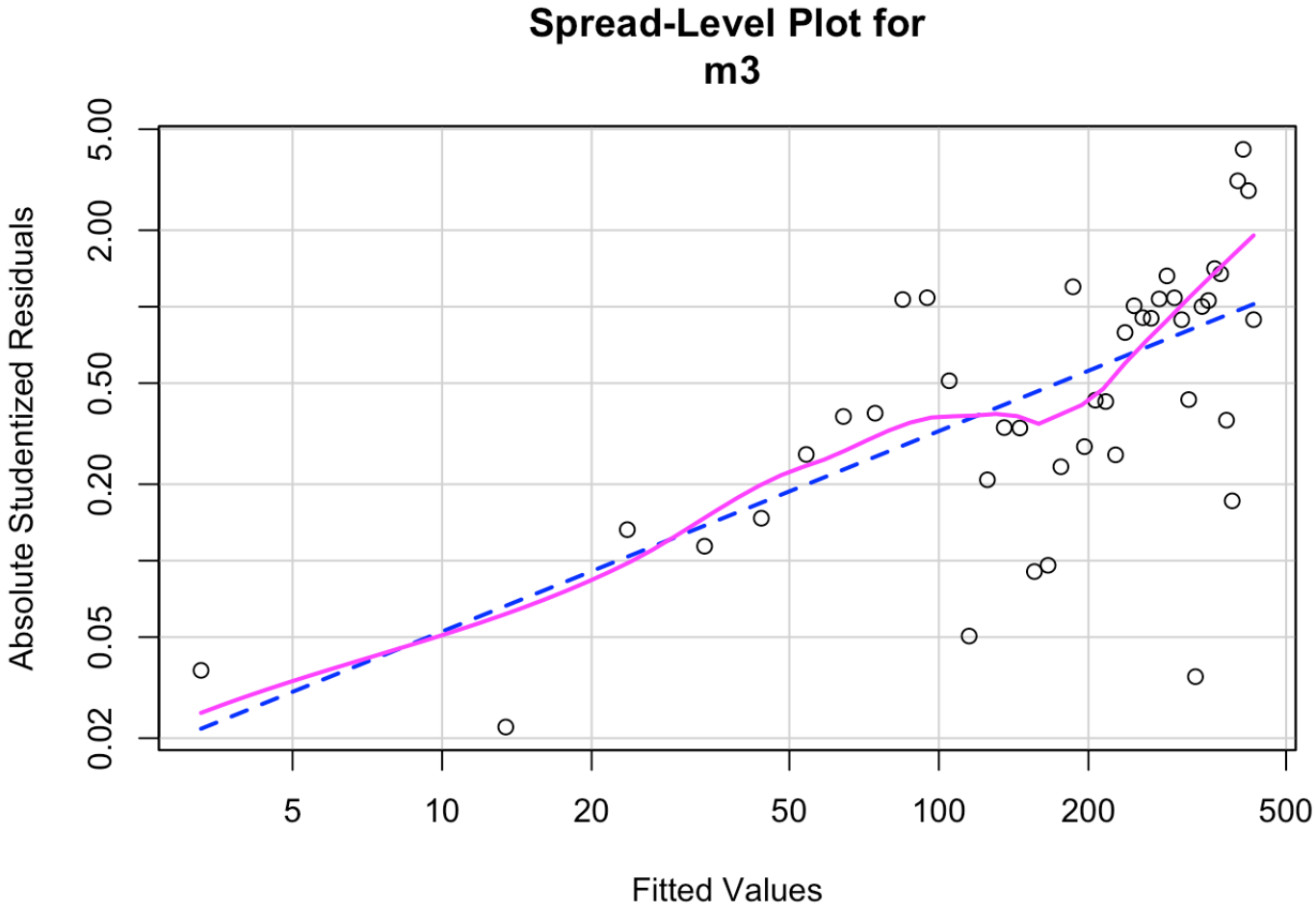
```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 29.20809, Df = 1, p = 6.5007e-08
```

```
plot(m3$fitted.values, m3$residuals)
```



```
# Spread Level Plot
myspread <- spreadLevelPlot(m3)
```

```
## Warning in spreadLevelPlot.lm(m3):
## 5 negative fitted values removed
```




```
myspread
```

```
##
## Suggested power transformation:  0.2107483
```

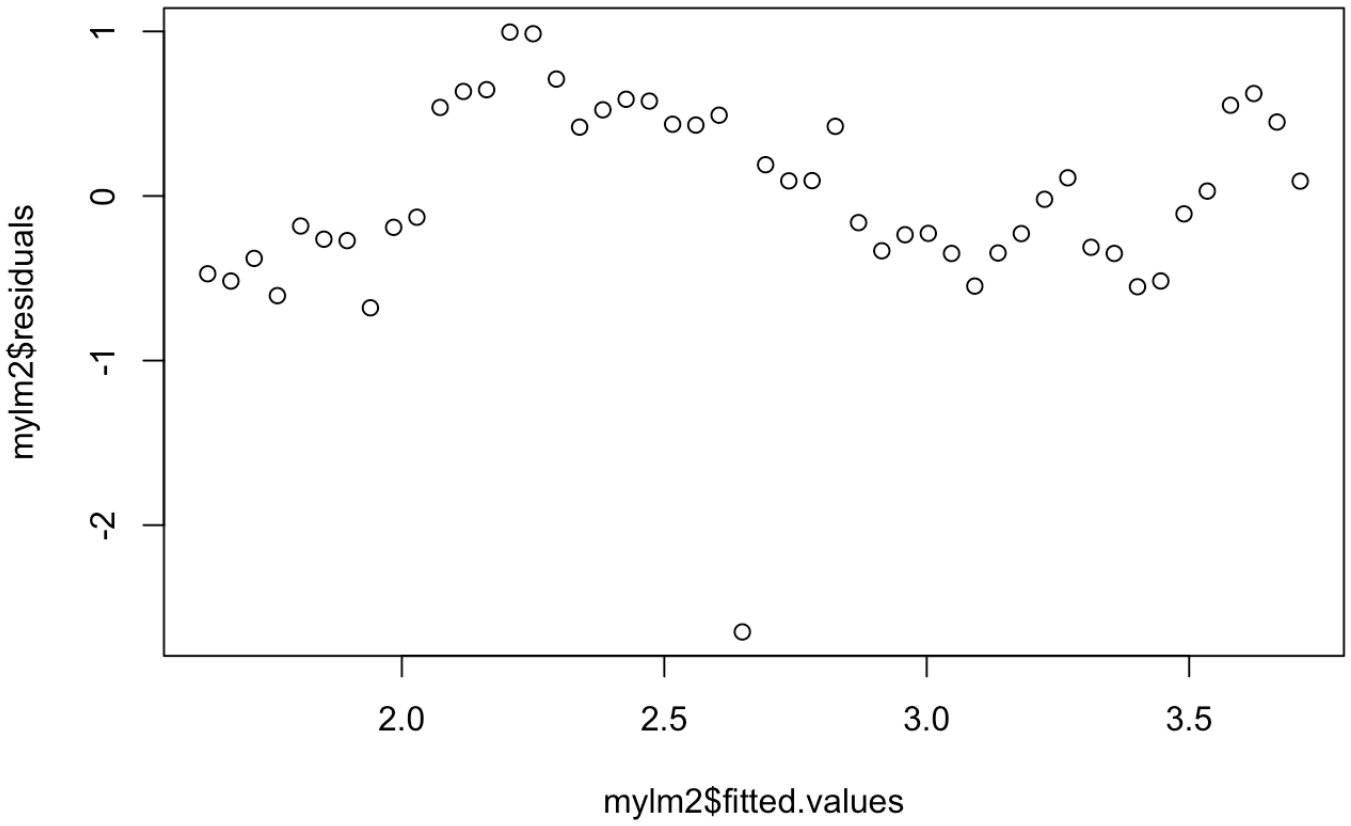
```
y <- df3_num$attacks
x <- df3_num$year

# Spread Level transformation and New Model
z<-y^(myspread$PowerTransformation)
mylm2<-lm(z ~ x)
summary(mylm2)
```

```
##
## Call:
## lm(formula = z ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6487 -0.3364 -0.0645  0.4596  0.9958
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -85.621139  12.558213  -6.818 1.71e-08 ***
## x             0.044290   0.006299   7.031 8.19e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6046 on 46 degrees of freedom
## Multiple R-squared:  0.518,  Adjusted R-squared:  0.5075
## F-statistic: 49.43 on 1 and 46 DF,  p-value: 8.193e-09
```

```
# Cook Weinberg Test Again
ncvTest(mylm2)                      # a high value of p indicates constant variance
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.7836975, Df = 1, p = 0.37601
```

```
plot(mylm2$fitted.values, mylm2$residuals)
```




```
# Normality Test
residuals2 <- mylm2$residuals
n <- 47
sd1 <- sd(mylm2$residuals)

ks.test(rnorm(n,0,sd1),mylm2$residuals)   # Large value of p indicates that residuals
are normal
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  rnorm(n, 0, sd1) and mylm2$residuals
## D = 0.2141, p-value = 0.1972
## alternative hypothesis: two-sided
```