

Global Terrorism Database - analiza

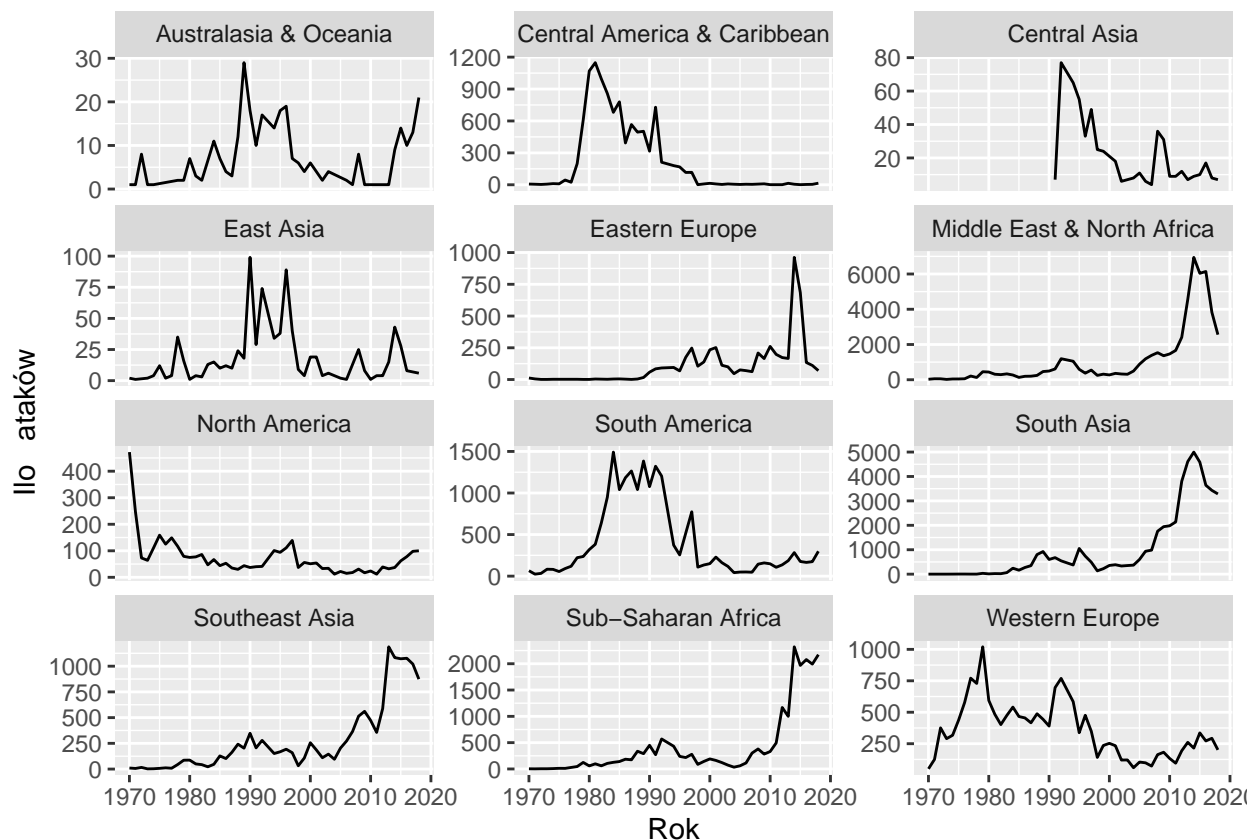
1. Wstęp

Raport zawiera analizę danych pochodzących z Global Terrorism Database (GTD) - bazy utrzymywanej przez amerykańskie National Consortium for the Study of Terrorism and Responses to Terrorism (START) na uniwersytecie Marylandu w College Park, zawierającej listę ponad 190 tys. ataków terrorystycznych od 1970 r. do końca 2018 r. (pomijając rok 1993). Każdy opisany jest przez najwyżej 135 zmiennych, określających m.in datę ataku, jego lokalizację, cel, rodzaj użytej broni lub narzędzi czy ilość poszkodowanych. Pełen opis bazy znajduje się w dołączonym dokumencie 'Codebook: Inclusion Criteria and Variables'.

2. Analiza wstępna/eksploracyjna

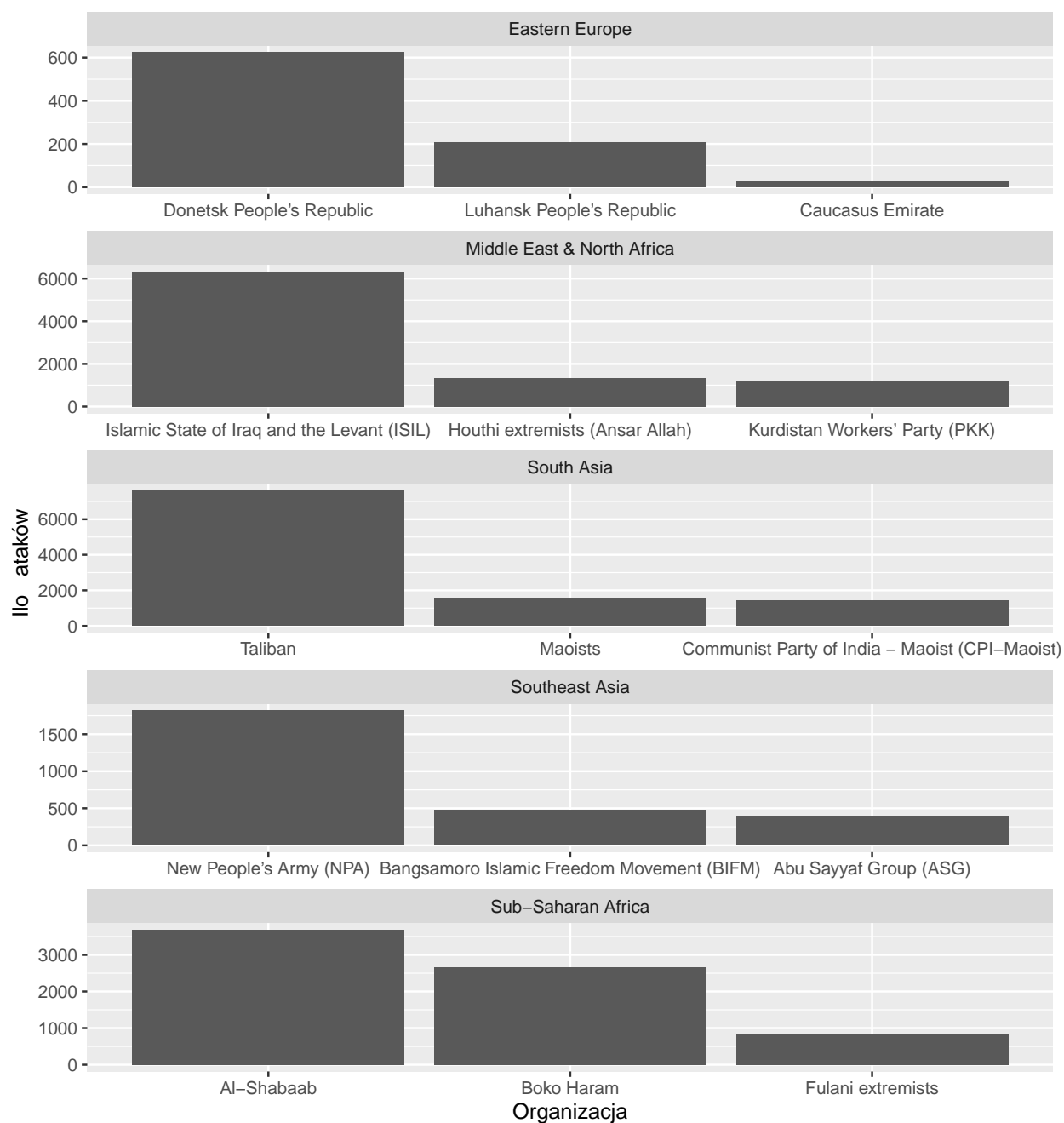
2.1 Ilość ataków rocznie wg regionu

Najpierw sprawdźmy, jak ilość ataków jest uzależniona od regionu świata, pozwoli nam to stwierdzić, które miejsca są obecnie najbardziej zagrożone atakiem terrorystycznym.



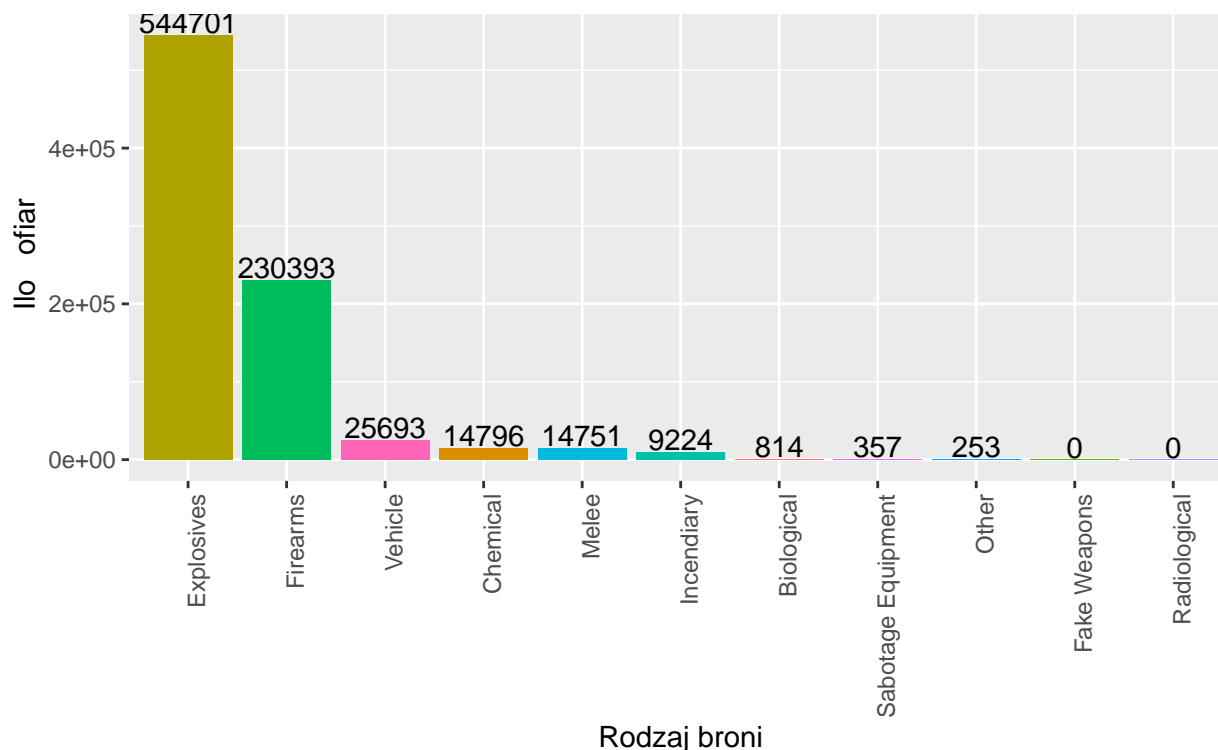
Na wykresie zauważyć można nagły wzrost ilości ataków na Bliskim Wschodzie i Afryce północnej, Afryce subsaharyjskiej oraz Azji południowej ok. roku 2014 - jest to najprawdopodobniej efekt wzrostu aktywności ugrupowań islamistycznych, zwłaszcza tzw. Państwa Islamskiego; widać także chwilowy wzrost ilości ataków w Europie wschodniej, co prawdopodobnie związane jest z pojawieniem się prorosyjskich ugrupowań separatystycznych na terenach wschodniej Ukrainy. W pozostałych regionach albo obserwujemy spadek, albo też ilość ataków terrorystycznych nie jest obecnie duża.

2.2 Najaktywniejsze ugrupowania terrorystyczne w Europie Wschodniej oraz na Bliskim Wschodzie, Afryce północnej i subsaharyjskiej, i Azji południowej i połudnowo-wschodniej od roku 2010



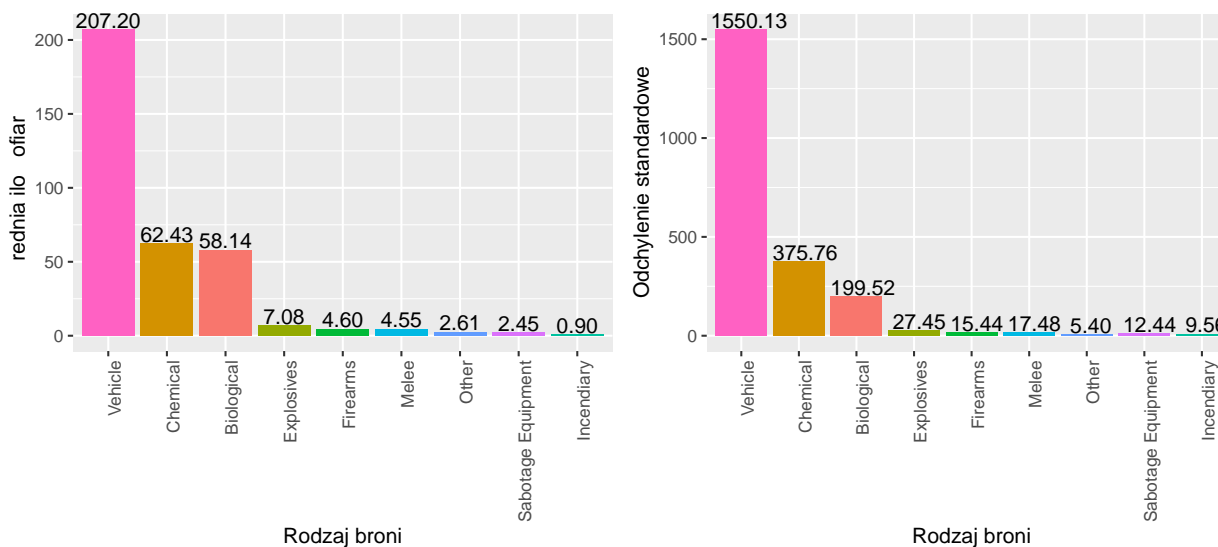
Wykres potwierdza przypuszczenia co do źródła ataków, pierwsze miejsca zajmują organizacje islamistyczne, w Europie wschodniej separatyści, natomiast w Azji południowo-wschodniej najwięcej ataków zorganizowała komunistyczna New People's Army z Filipin, jednak drugie i trzecie miejsce przypada organizacjom islamistycznym.

2.3 Sumaryczna ilość ofiar (zabitych lub rannych) w zależności od zastosowanej broni



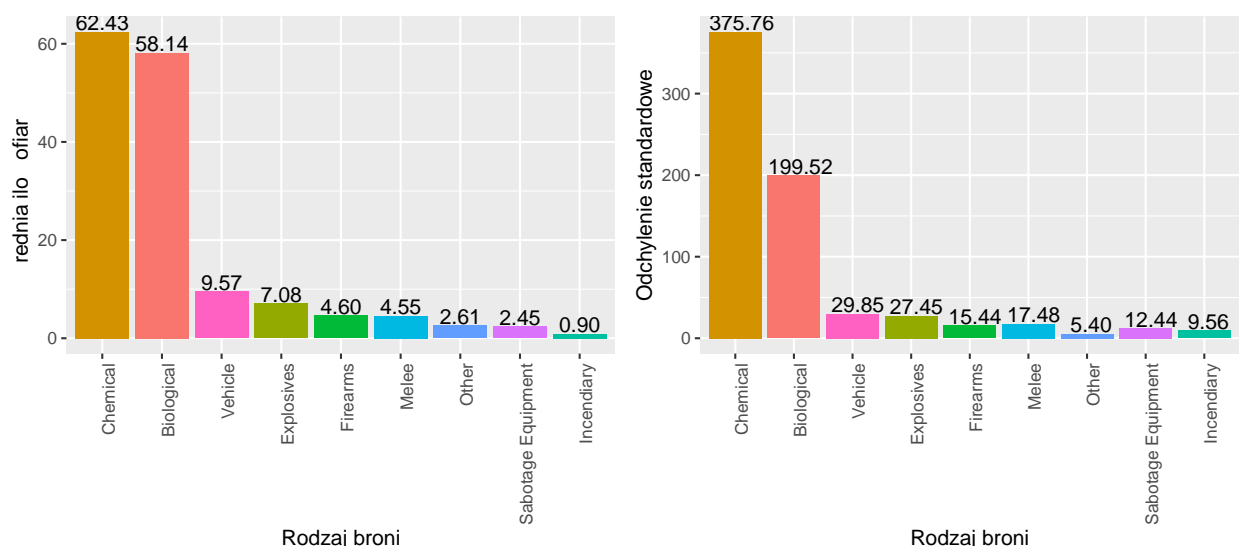
Wzięto pod uwagę jedynie udane ataki (parametr *success* równy 1), których liczba ofiar i rodzaj użytego uzbrojenia są znane. Największą liczbę ofiar spowodowały ładunki wybuchowe - ponad 2 razy więcej niż reszta rodzajów broni razem wzięta, co może świadczyć zarówno o skuteczności tego narzędzia jak i o popularności. Żeby to sprawdzić, należy policzyć średnią ilość ofiar na atak, a także odchylenie standardowe, co pozwoli stwierdzić jak bardzo różne efekty daje zastosowanie danego rodzaju broni.

2.4 Średnia i odchylenie standardowe ilości ofiar w zależności od zastosowanej broni



Wartość średniej i odchylenia standardowego dla kategorii *Vehicle* sugeruje występowanie wartości odstających

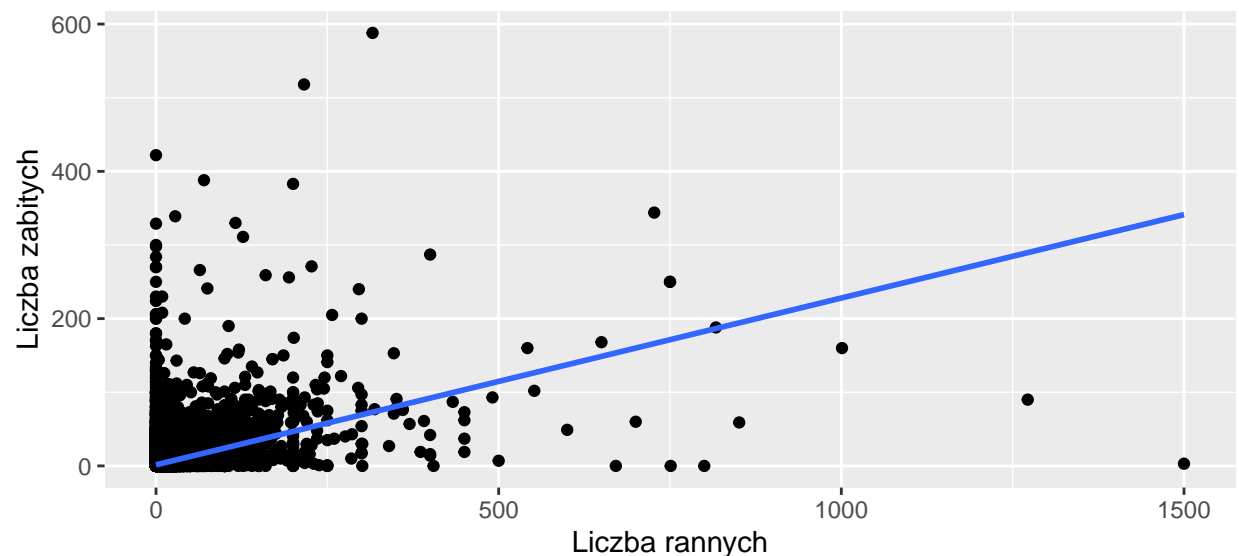
w danych, zawyżających wynik na tyle, że uzyskane wartości nie pozwalają na jednoznaczne określenie relacji pomiędzy poziomami skuteczności różnych rodzajów uzbrojenia. Po przejrzeniu danych udało się znaleźć przyczynę takiego stanu rzeczy - jest to uwzględnienie informacji nt. ataku na WTC z 11.09.2001 r. Po odrzuceniu odpowiadających mu wpisów otrzymujemy:



Teraz widać, że najwyższą średnią ofiar przynoszą ataki z zastosowaniem broni chemicznej lub biologicznej, ataki przy użyciu materiałów wybuchowych są dopiero na czwartym miejscu, co oznacza, że wysoka łączna ilość ofiar ataków z zastosowaniem materiałów wybuchowych wynika jedynie z popularności tej metody ataku.

3. Modelowanie statystyczne

3.1 Zależność między ilością zabitych w ataku a ilością rannych

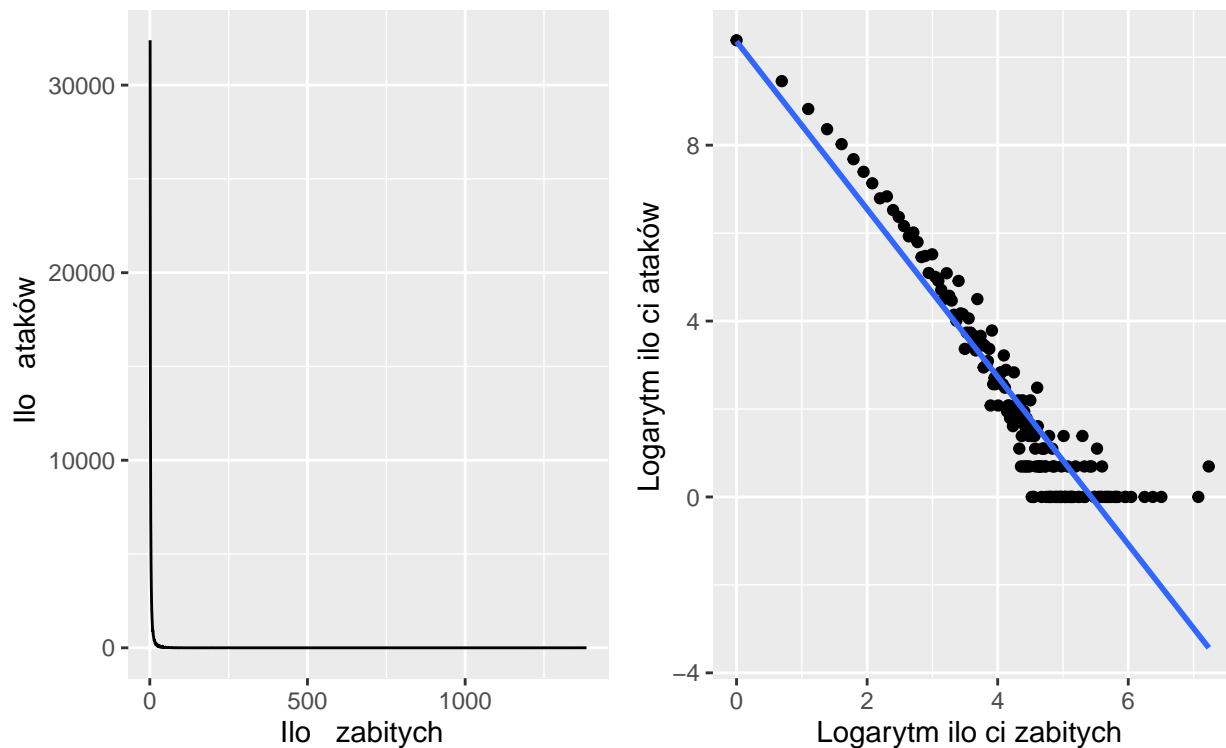


Zauważyć można wzrost liczby zabitych przy rosnącej liczbie rannych - wytłumaczyć to można wzrostem skali ataku. Parametry prostej regresji:

```
##
## Call:
## lm(formula = nkill ~ nwound, data = tmp)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -338.25   -1.49    -1.49    -0.49   514.94
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.486323   0.019674   75.55  <2e-16 ***
## nbound       0.226512   0.001254  180.68  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.209 on 141040 degrees of freedom
## Multiple R-squared:  0.188, Adjusted R-squared:  0.188
## F-statistic: 3.265e+04 on 1 and 141040 DF, p-value: < 2.2e-16
```

3.2 Analiza rozkładu ilości zabitych w ataku terrorystycznym



Wykresy przedstawiają ilość ataków w zależności od liczby zabitych, przy czym wzięto pod uwagę jedynie te o niezerowej liczbie ofiar. Na drugim wykresie widać, że zależność między logarytmem ilości ataków a logarytmem ilości zabitych jest bliska liniowej, co sugeruje, że rozkład ilości zabitych w ataku terrorystycznym jest zgodny z prawem potęgowym (power law), czyli, że funkcja prawdopodobieństwa rozkładu ma postać $p(x) = P(X = x) = Cx^{-\alpha}$. W celu sprawdzenia dopasowania rozkładu wykorzystano pakiet *poweRlaw* i wchodzącej w jego skład funkcji *bootstrap_p*. Uzyskana w ramach testu wartość p :

```
## Expected total run time for 100 sims, using 2 threads is 92.8 seconds.
```

```
## [1] 0.33
```

pozwala na przyjęcie hipotezy zerowej. Estymowana wartość parametru α wynosi:

```
## [1] 1.782875
```