

# ADPS 2025Z — Laboratorium 1

Konrad Jędrzejewski, Marek Rupniewski

## Przykład 1

### Pobranie danych

- W katalogu domowym utwórz katalog ADPS/Lab1. W tym celu w RStudio w prawym dolnym oknie w zakładce *Files* wybierz *New Folder*.

W RStudio zmień katalog roboczy na ADPS/Lab1. W zakładce *Files* wybierz *More -> Set As Working Directory*

- Znajdź i pobierz dane historyczne spółki KGHM (KGH) ze strony <https://stooq.pl/>:

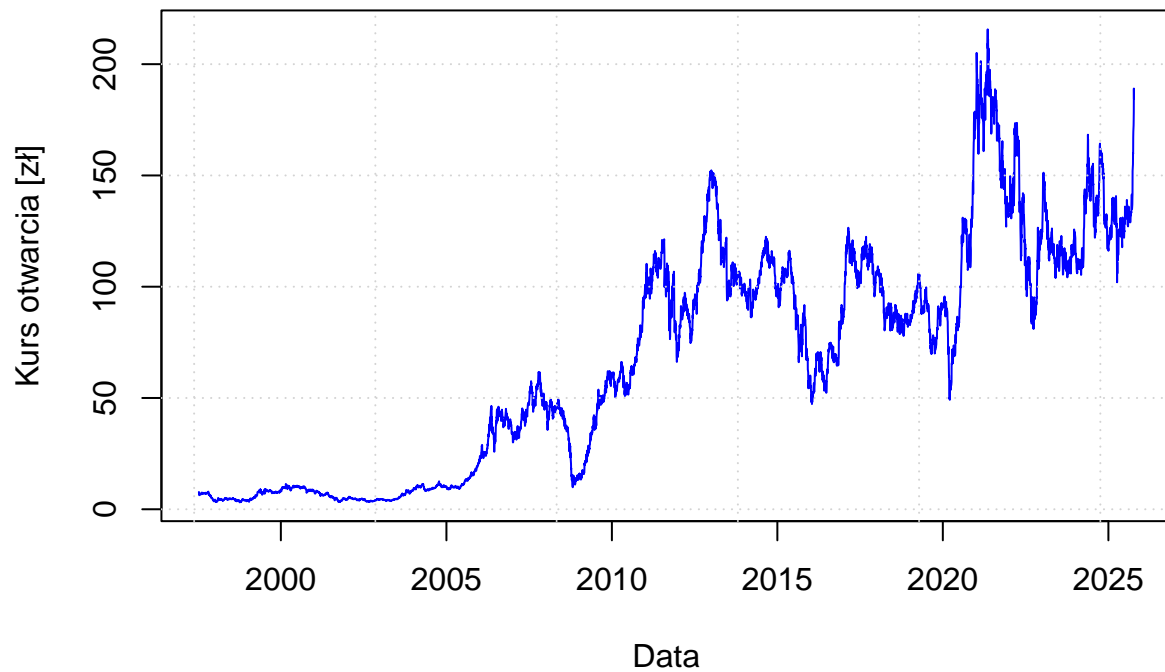
```
# https://stooq.pl/q/d/l/?s=kghe&i=d
Ticket = 'KGH'
webLink = paste0('https://stooq.pl/q/d/l/?s=', Ticket, '&i=d')
fileName = paste0(Ticket, '.csv')
# if(!file.exists(fileName)) {
  download.file(webLink, fileName)
#}
```

### Wykres kursu w czasie

- Wczytaj dane z pliku .csv. Obejrzyj dane dotyczące spółki KGHM, zwróć uwagę na nazwy kolumn.
- Narysuj wykres kursu otwarcia w zależności od daty:

```
df_KGH = read.csv('KGH.csv')
df_KGH$Data = as.Date(df_KGH$Data)
plot(Otwarcie ~ Data, df_KGH, type = 'l', col = 'blue',
     xlab = 'Data', ylab = 'Kurs otwarcia [zł]', main = 'KGHM' )
grid()
```

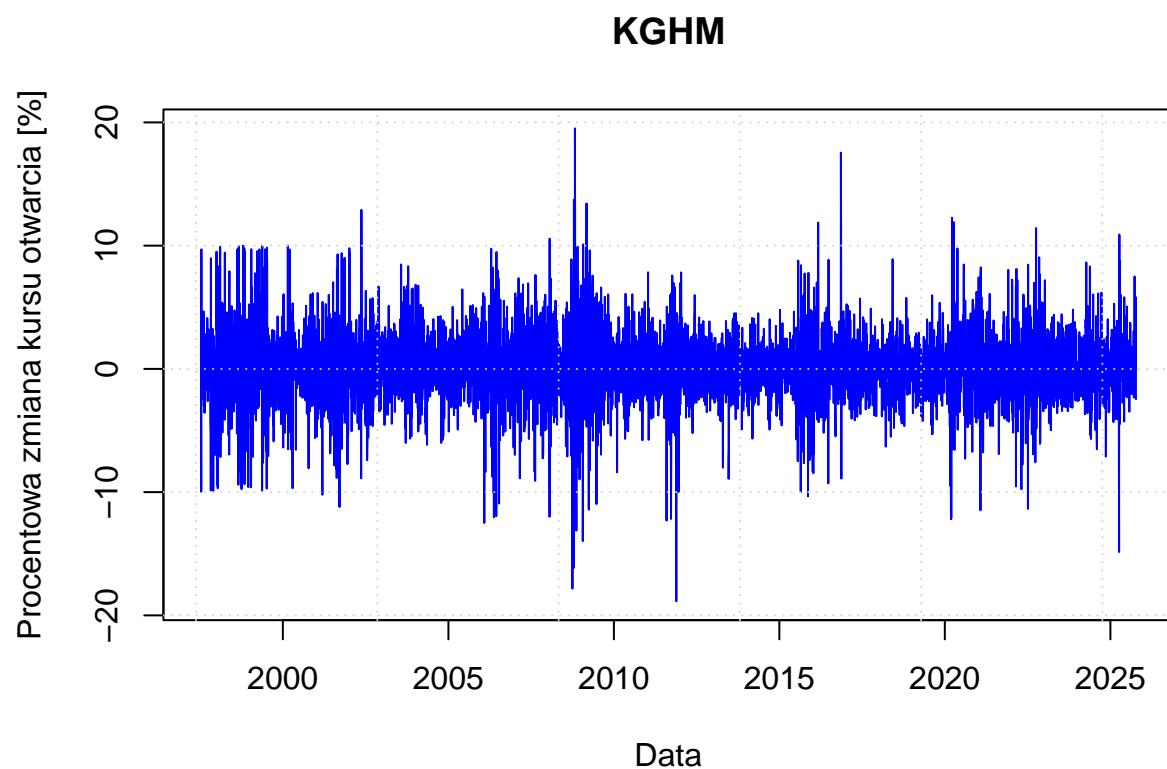
## KGHM



### Procentowe zmiany kursu

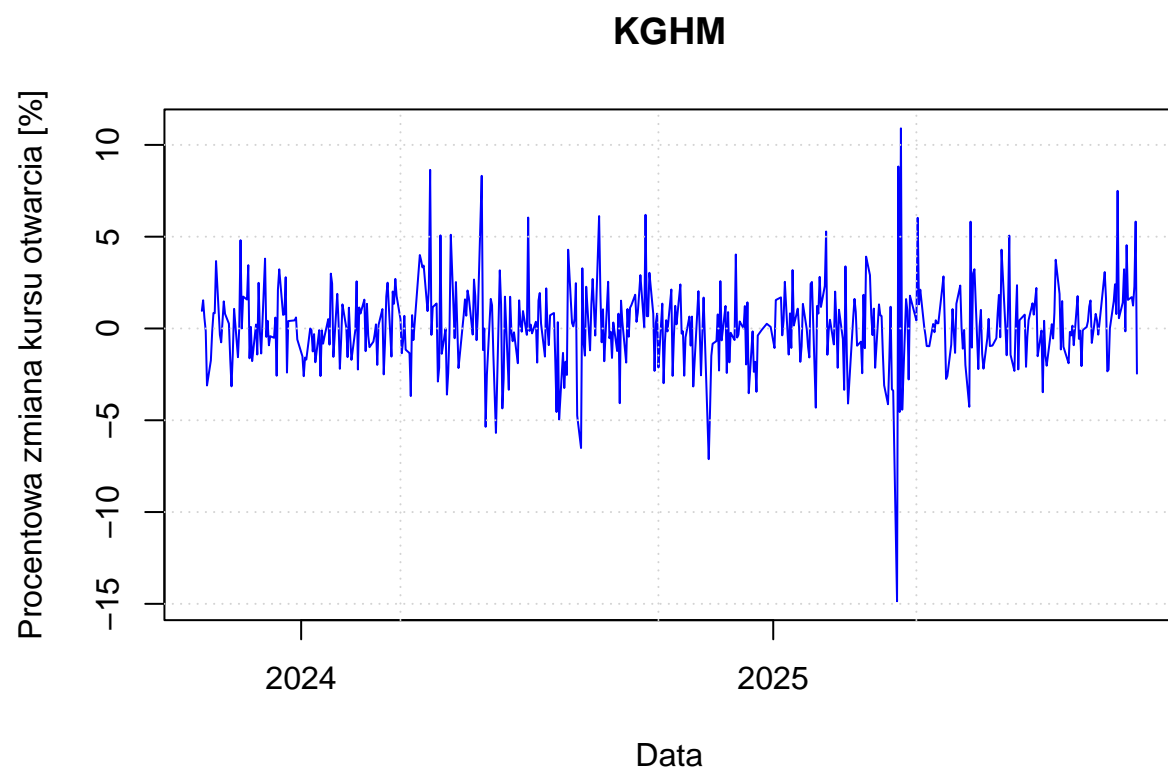
- Wykres procentowych zmian kursu otwarcia:

```
df_KGH$Otwarcie_zm = with(df_KGH, c(NA, 100*diff(Otwarcie)/Otwarcie[-length(Otwarcie)]))
plot(Otwarcie_zm ~ Data, df_KGH, type = 'l', col = 'blue', xlab = 'Data',
     ylab = 'Procentowa zmiana kursu otwarcia [%]', main = 'KGHM')
grid()
```



Ograniczenie danych do danych z ostatnich dwóch lat

```
df_KGH_2lata = df_KGH[which(df_KGH$Data >= '2023-10-16' & df_KGH$Data <= Sys.Date()),]  
plot(Otwarcie_zm ~ Data, df_KGH_2lata, type = 'l', col = 'blue', xlab = 'Data',  
      ylab = 'Procentowa zmiana kursu otwarcia [%]', main = 'KGHM')  
grid()
```

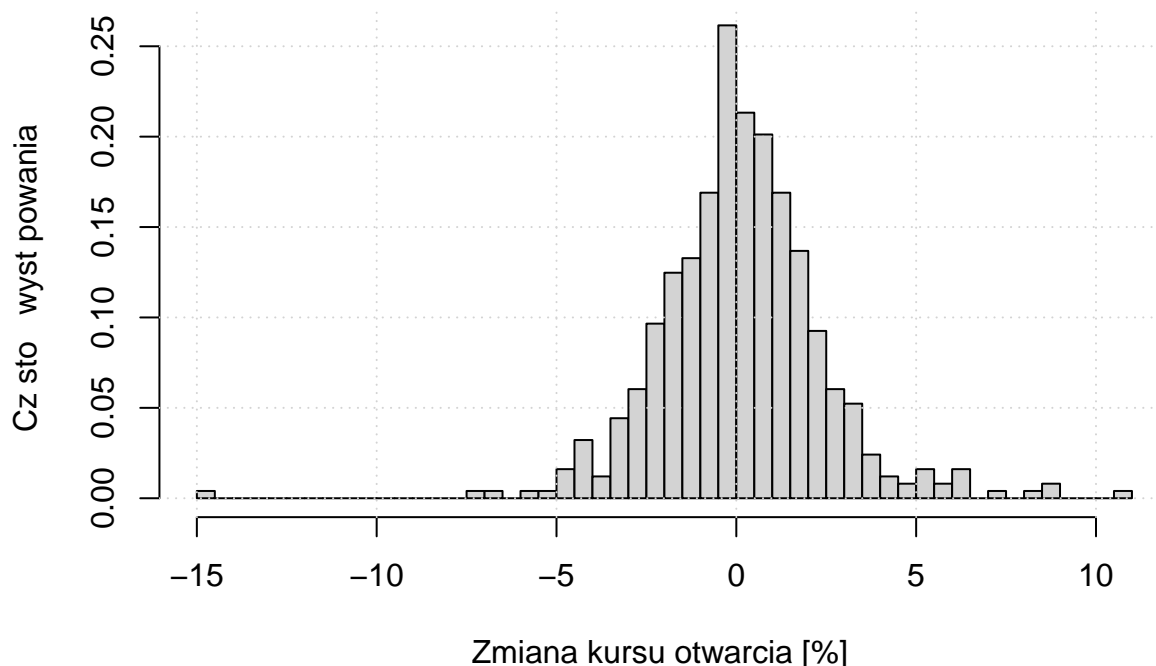


## Histogram

- Histogram procentowych zmian kursu otwarcia:

```
hist(df_KGH_2lata$Otwarcie_zm, breaks = 50, prob = T,  
     xlab = 'Zmiana kursu otwarcia [%] ',  
     ylab = 'Częstość występowania',  
     main = paste('Histogram procentowych zmian kursu', 'KGHM') )  
grid()
```

## Histogram procentowych zmian kursu KGHM



### Funkcja gęstości prawdopodobieństwa

- Wartość średnia oraz odchylenie standardowe zmian kursu otwarcia:

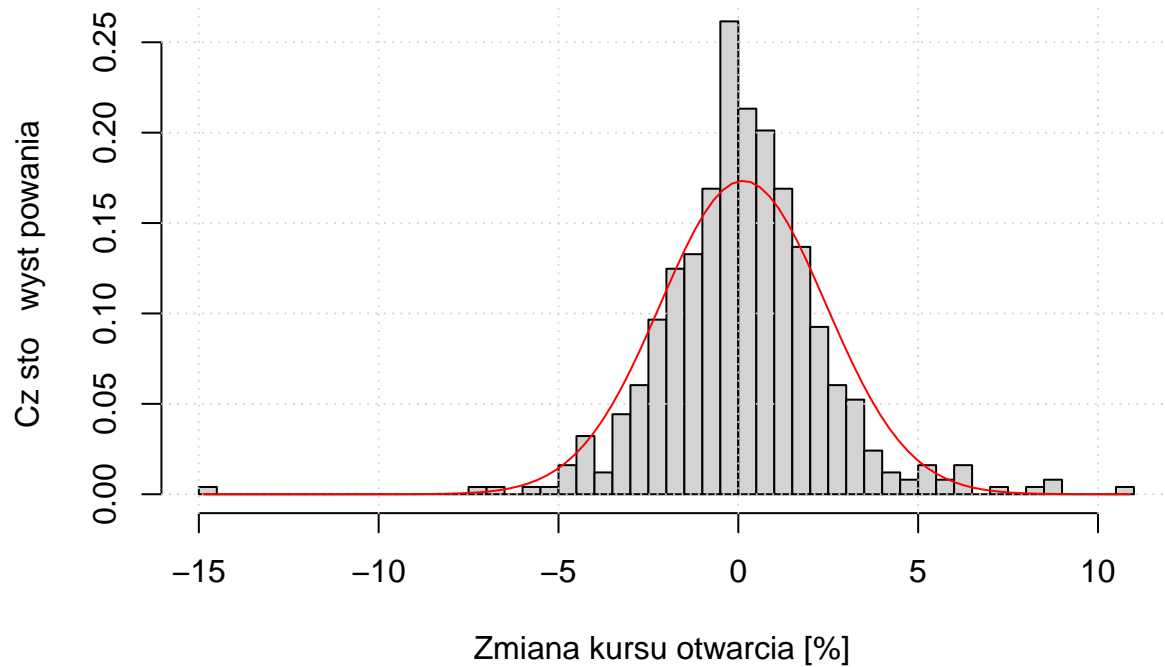
```
m = mean(df_KGH_2lata$Otwarcie_zm, na.rm = T)
s = sd(df_KGH_2lata$Otwarcie_zm, na.rm = T)
```

Wartość średnia zmian kursu otwarcia wynosi 0.1335 %, a odchylenie standardowe 2.3012 %.

- Dorysowanie do histogramu wykres gęstości rozkładu normalnego o obliczonych parametrach:

```
hist(df_KGH_2lata$Otwarcie_zm, breaks = 50, prob = T,
     xlab = 'Zmiana kursu otwarcia [%] ',
     ylab = 'Częstość występowania',
     main = paste('Histogram procentowych zmian kursu', 'KGHM') )
grid()
min_c = min(df_KGH_2lata$Otwarcie_zm, na.rm = T)
max_c = max(df_KGH_2lata$Otwarcie_zm, na.rm = T)
curve(dnorm(x, mean = m, sd = s), add = T, col = 'red', from = min_c, to = max_c)
```

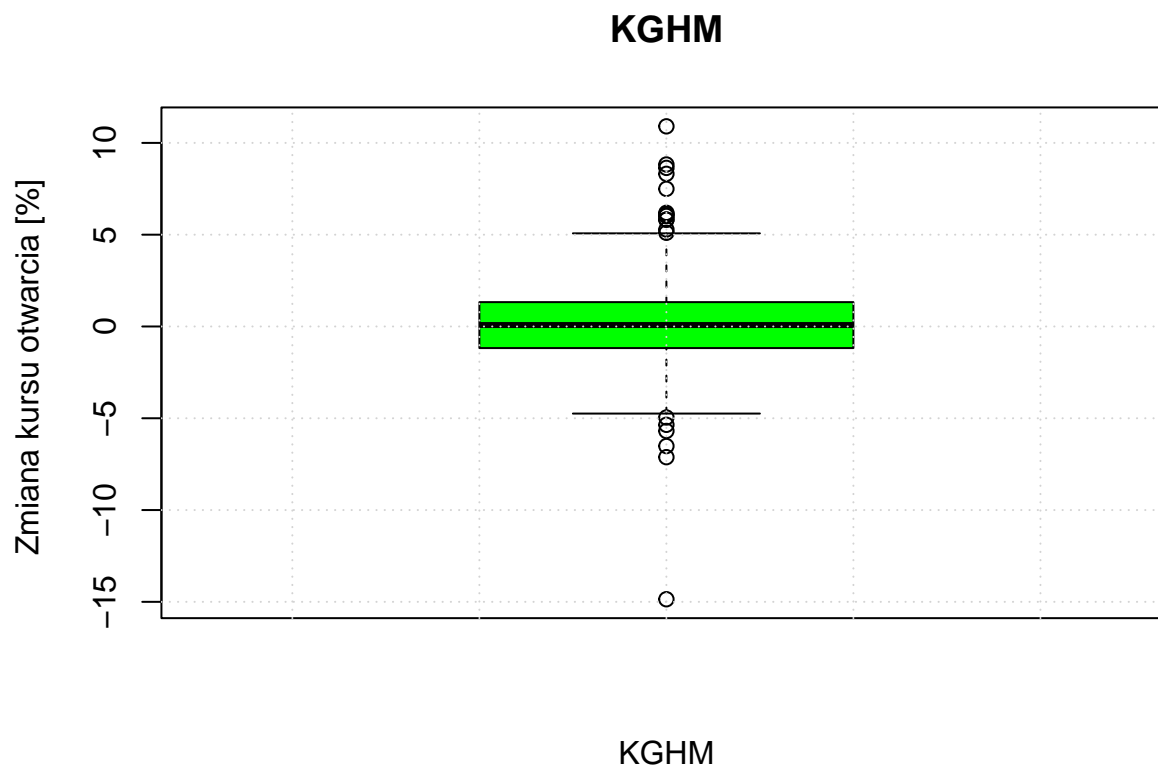
## Histogram procentowych zmian kursu KGHM



### Wykres pudełkowy

- Wykres pudełkowy procentowych zmian kursu otwarcia:

```
boxplot(df_KGH_2lata$Otwarcie_zm, col = 'green',  
        xlab = 'KGHM', ylab = 'Zmiana kursu otwarcia [%] ',  
        main = 'KGHM' )  
grid()
```



## Przykład 2

### Pobranie danych

- Wgraj dane historyczne dotyczące katastrof lotniczych znajdujące się w pliku `crashes.csv` do katalogu `ADPS/Lab1`.
- Wczytaj dane z pliku `crashes.csv` do środowiska R:

```
kat = read.csv('crashes.csv')
```

- Obejrzyj dane, zwróć uwagę na puste pola.

### Liczba wypadków w latach

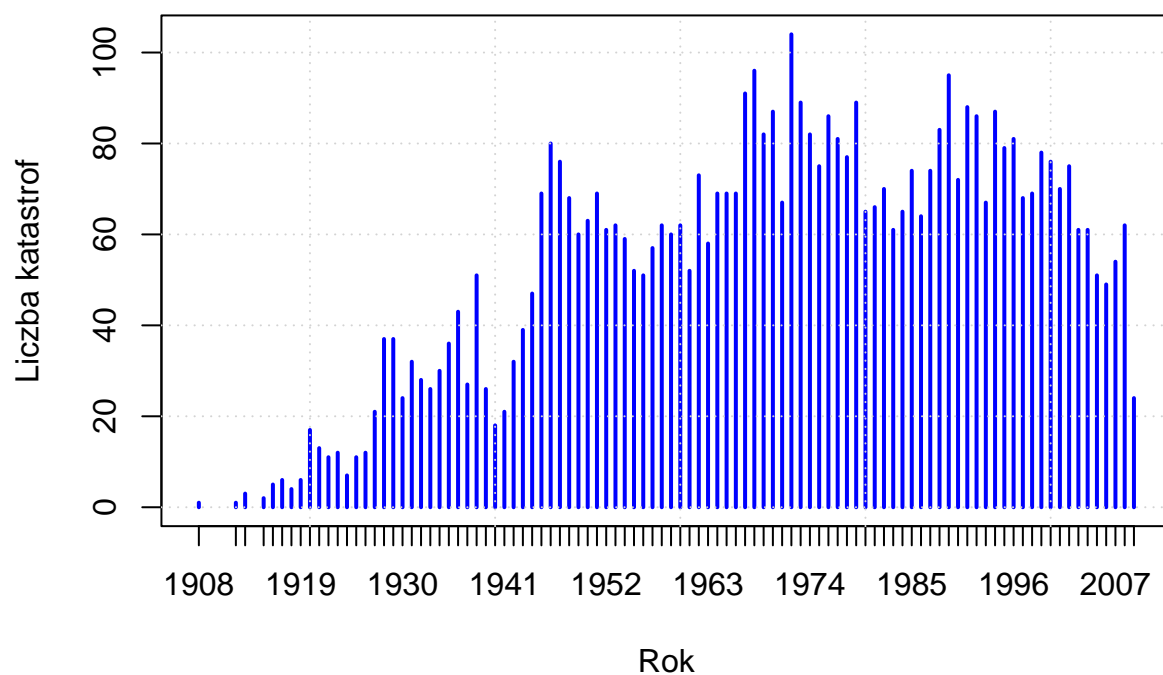
- Dodanie do danych kolumny z rokiem:

```
kat$Year = strftime(as.Date(kat$Date, '%m/%d/%Y'), '%Y')
```

- Wykres liczby wypadków w danym roku:

```
plot(table(kat$Year), type = 'h', col = 'blue', xlab = 'Rok',
      ylab = 'Liczba katastrof', main = 'Liczba katastrof w roku' )
grid()
```

## Liczba katastrof w roku



## Liczba ofiar w latach

- Agregacja danych po latach:

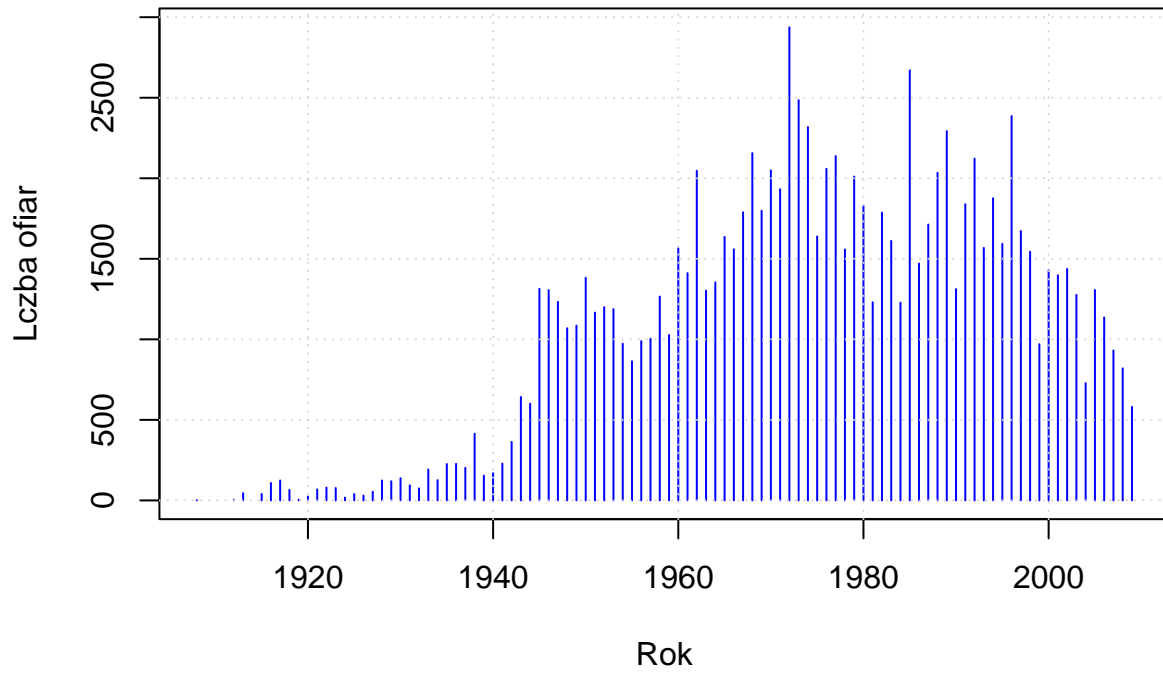
```
Ofiary_agr = aggregate(Fatalities ~ Year, kat, FUN = sum)
```

- Wykres:

```
plot(Ofiary_agr, type = 'h', col = 'blue', xlab = 'Rok',  
     ylab = 'Liczba ofiar', main = 'Liczba ofiar katastrof w roku' )  
grid()
```



## Liczba ofiar katastrof w roku

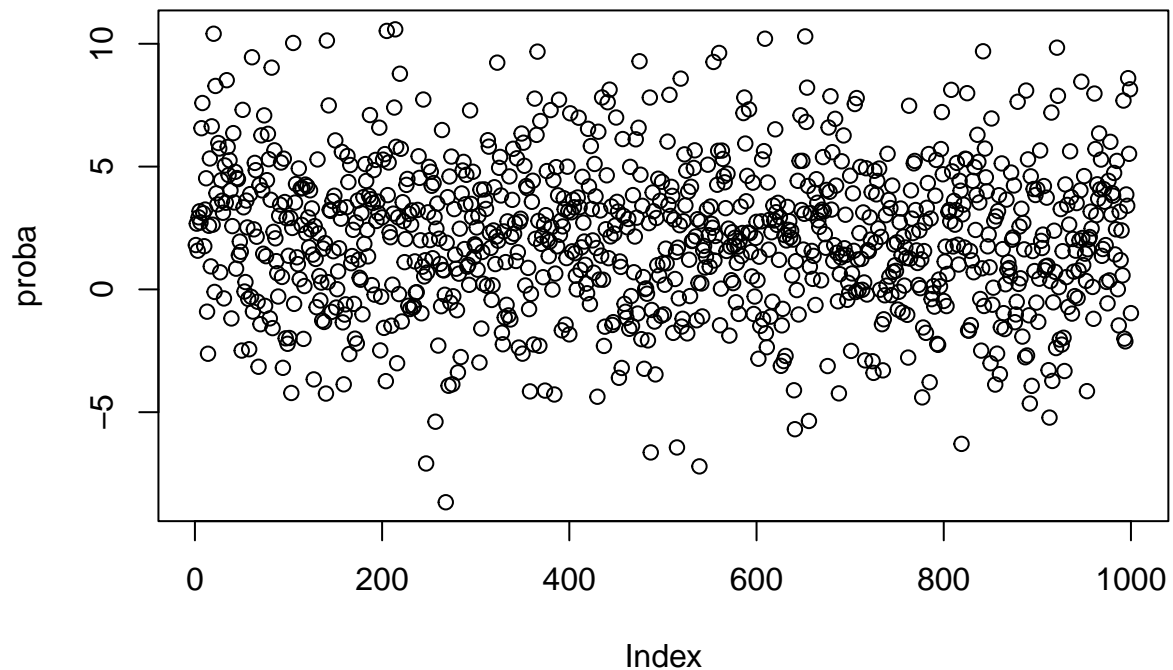


---

## Przykład 3

- Generacja 1000 próbek z rozkładu normalnego  $N(2,9)$ :

```
proba = rnorm(1000, mean = 2, sd = 3)
plot(proba)
```



- Wartości parametrów z próby

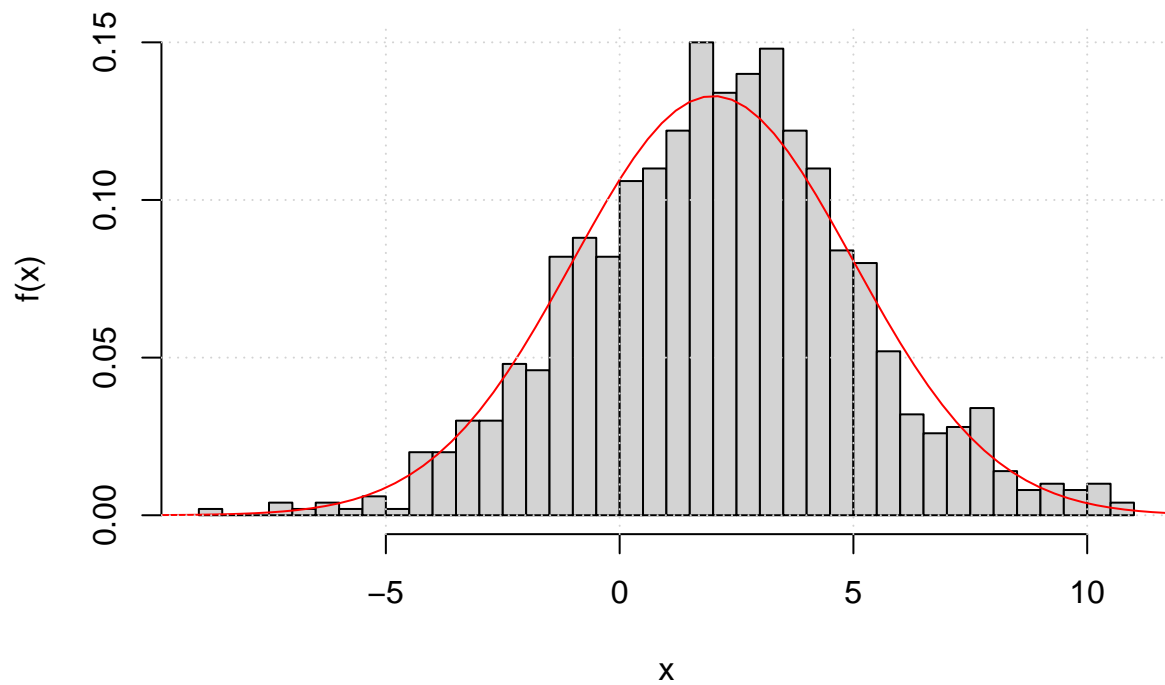
```
m = mean(proba); s = sd(proba)
```

Wystymowane parametry rozkładu wynoszą: wartość średnia 2.1073, wariancja 8.9339.

- Histogram i gęstość prawdopodobieństwa:

```
hist(proba, breaks = 50, prob = T, xlab = 'x', ylab = 'f(x)',  
     main = 'Histogram i gęstość prawdopodobieństwa')  
curve(dnorm(x, mean = 2, sd = 3), add = T, col = 'red', -15, 15)  
grid()
```

## Histogram i gsto prawdopodobieństwa

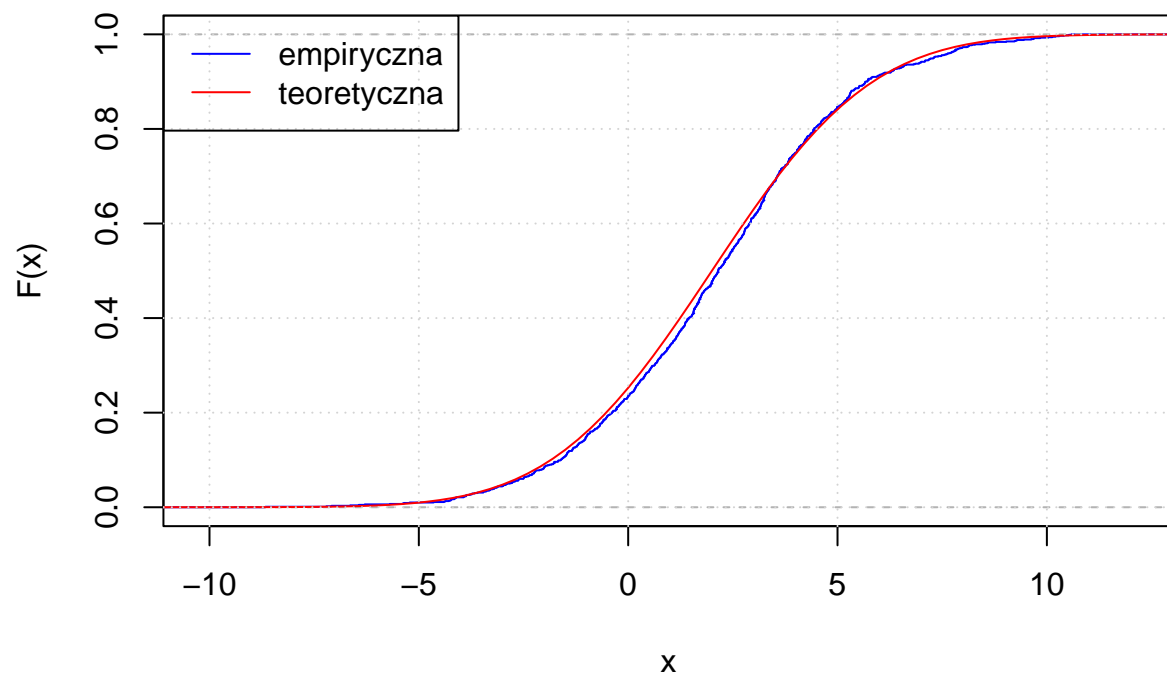


- Dystrybuanta empiryczna i teoretyczna:

```
plot(ecdf(proba), xlab = 'x', ylab = 'F(x)', main = 'Dystrybuanta', col = 'blue')
curve(pnorm(x, mean = 2, sd = 3), add = T, col = 'red', -15, 15)
grid()

legend('topleft', c('empiryczna', 'teoretyczna'),
      col = c('blue', 'red'), lwd = 1)
```

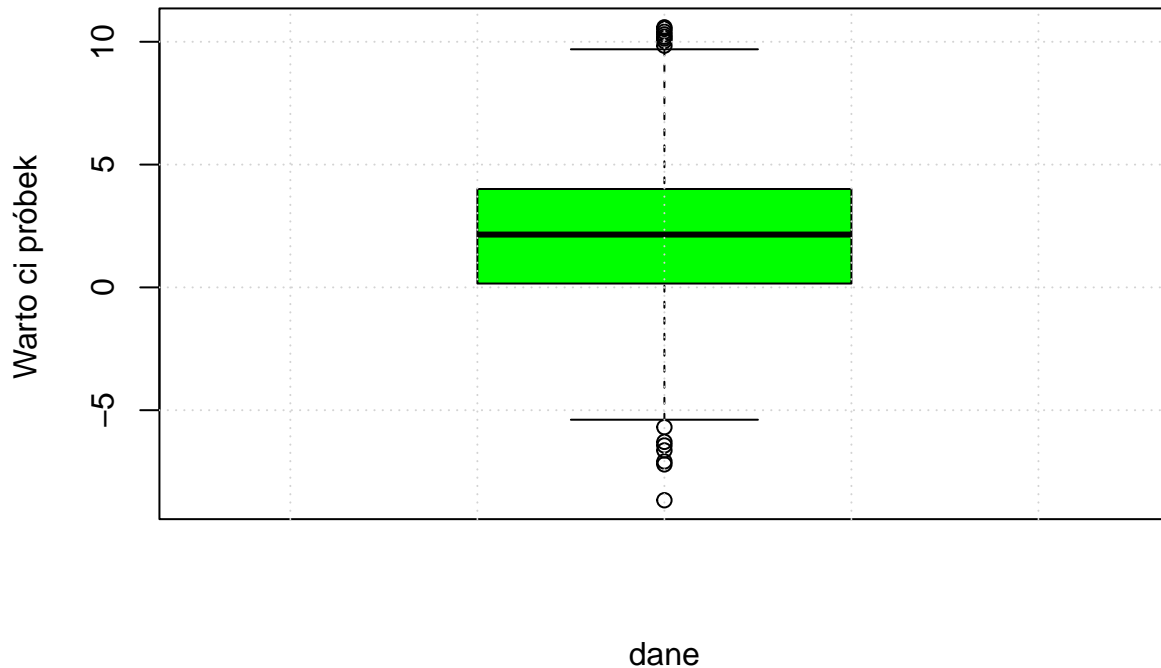
## Dystrybuanta



- Wykres pudełkowy:

```
boxplot(proba, col = 'green', xlab = 'dane', ylab = 'Wartości próbek', main = 'Wykres pudełkowy')  
grid()
```

## Wykres pudełkowy



- Teoretyczne i empiryczne wartości kwantyli dla 0.25, 0.5 i 0.75:

```
q_theor = qnorm(c(0.25, 0.5, 0.75), mean = 2, sd = 3)
q_empir = quantile(proba, c(0.25, 0.5, 0.75))
```

Kwartyle teoretyczne: -0.0235, 2, 4.0235.

Kwartyle empiryczne: 0.1584, 2.1503, 3.9998.

## Przykład 4

- Generacja 1000 próbek z rozkładu Poissona  $Pois(4)$ :

```
M = 1000
proba = rpois(M, lambda = 4)
```

- Wartości parametrów z próby:

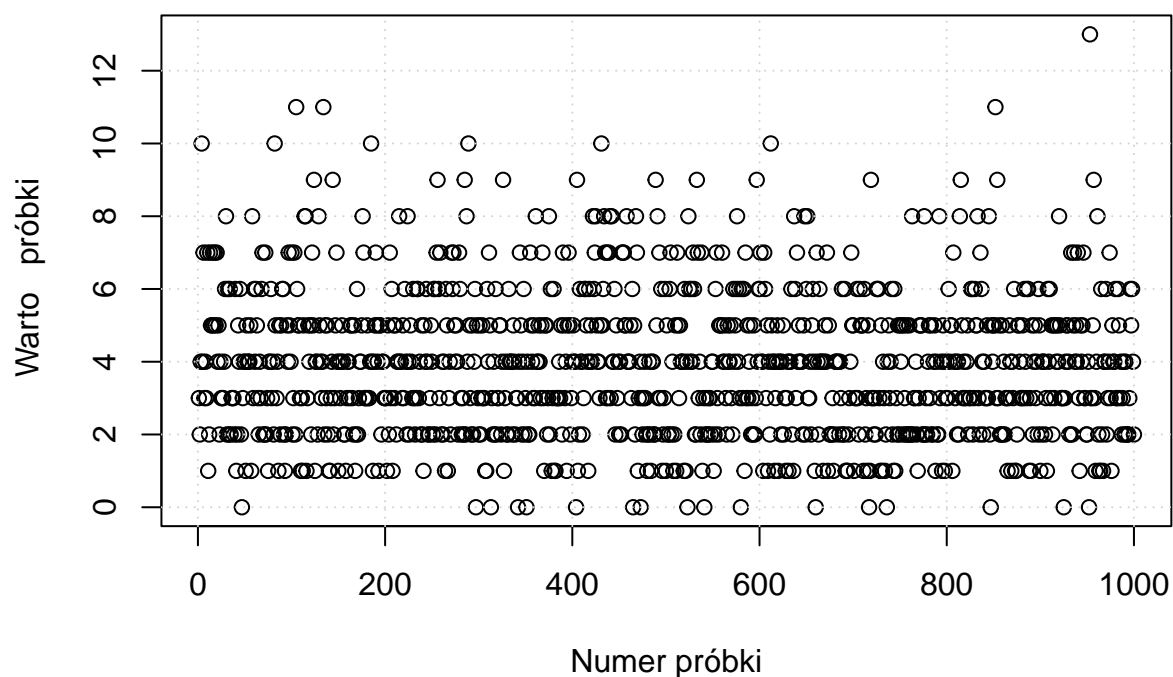
```
m = mean(proba); v = var(proba)
```

Wyestymowane parametry rozkładu wynoszą: wartość średnia 3.918, wariancja 4.2135.

- Wykres próbek:

```
plot(proba, xlab = 'Numer próbki', ylab = 'Wartość próbki',
     main = 'Wartości wygenerowanych próbek')
grid()
```

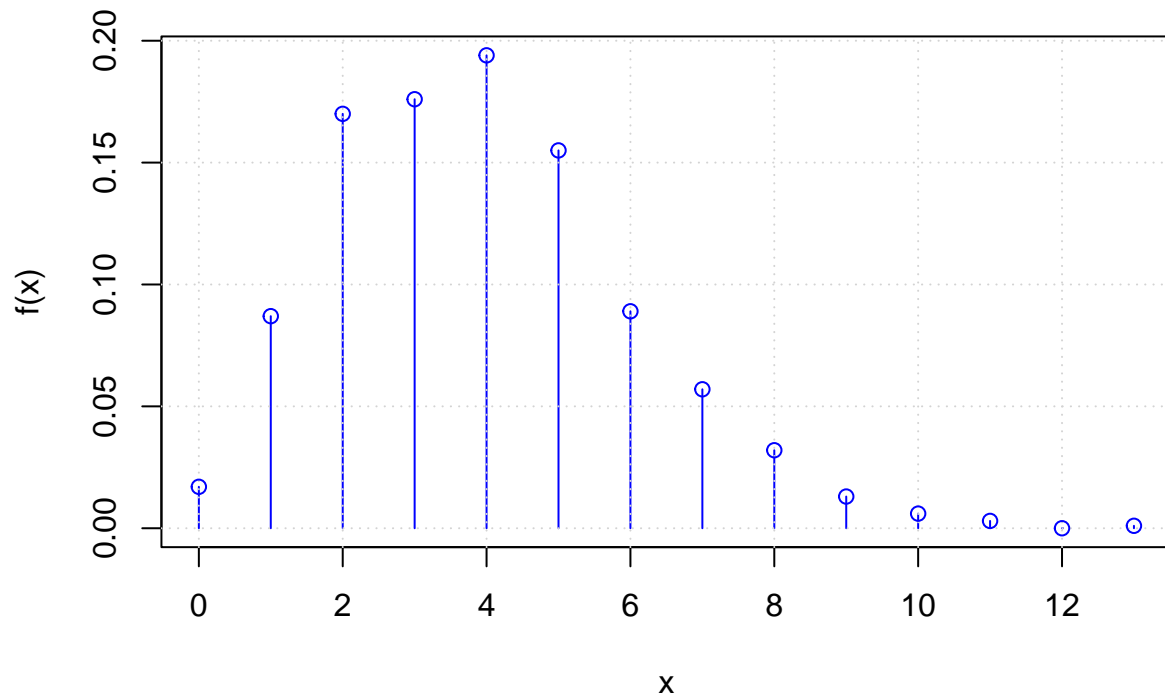
## Warto ci wygenerowanych próbek



- Empiryczna funkcja prawdopodobieństwa:

```
Arg = 0:max(proba)
Freq = as.numeric(table(factor(proba, levels = Arg))) / M
plot(Freq ~ Arg, type = 'h', col = 'blue', xlab = 'x', ylab = 'f(x)',
     main = paste0('Funkcja prawdopodobieństwa dla M = ', M))
grid()
points(Freq ~ Arg, col = 'blue')
```

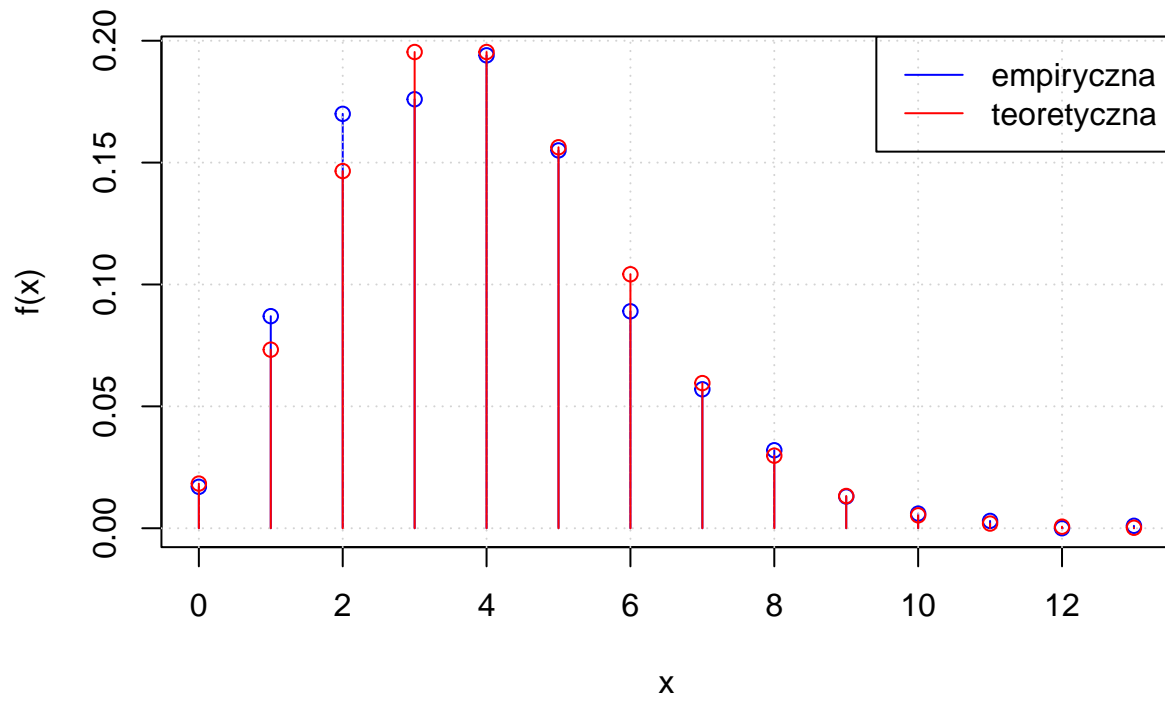
## Funkcja prawdopodobieństwa dla $M = 1000$



- Teoretyczna funkcja prawdopodobieństwa:

```
plot(Freq ~ Arg, type = 'h', col = 'blue', xlab = 'x', ylab = 'f(x)',  
     main = paste0('Funkcja prawdopodobieństwa dla M = ', M))  
grid()  
points(Freq ~ Arg, col = 'blue')  
  
lines(dpois(Arg, lambda = 4) ~ Arg, type = 'h', col = 'red',  
      xlab = 'x', ylab = 'f(x)')  
points(dpois(Arg, lambda = 4) ~ Arg, col = 'red')  
  
legend('topright', c('empiryczna', 'teoretyczna'),  
      col = c('blue', 'red'), lwd = 1)
```

## Funkcja prawdopodobieństwa dla $M = 1000$

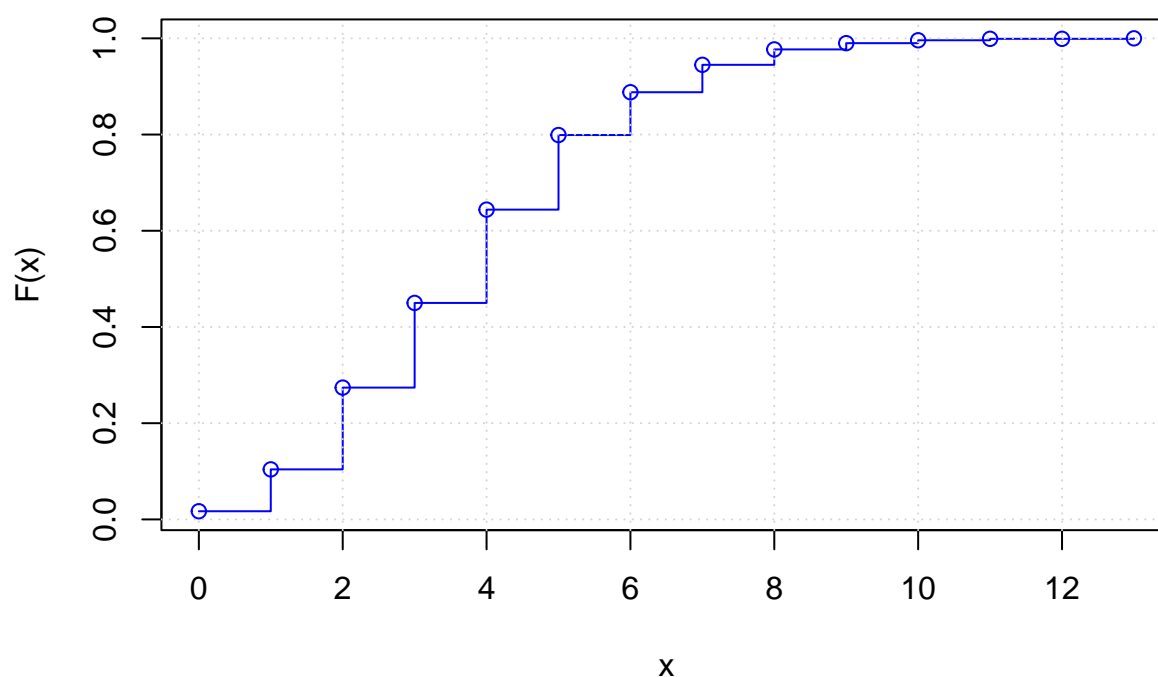


- Dystrybuanta empiryczna:

```
plot(cumsum(Freq) ~ Arg, type = 's', col = 'blue',
     xlab = 'x', ylab = 'F(x)', main = paste0('Dystrybuanta dla M = ', M))
grid()
points(cumsum(Freq) ~ Arg, col = 'blue')
```



## Dystrybuanta dla $M = 1000$



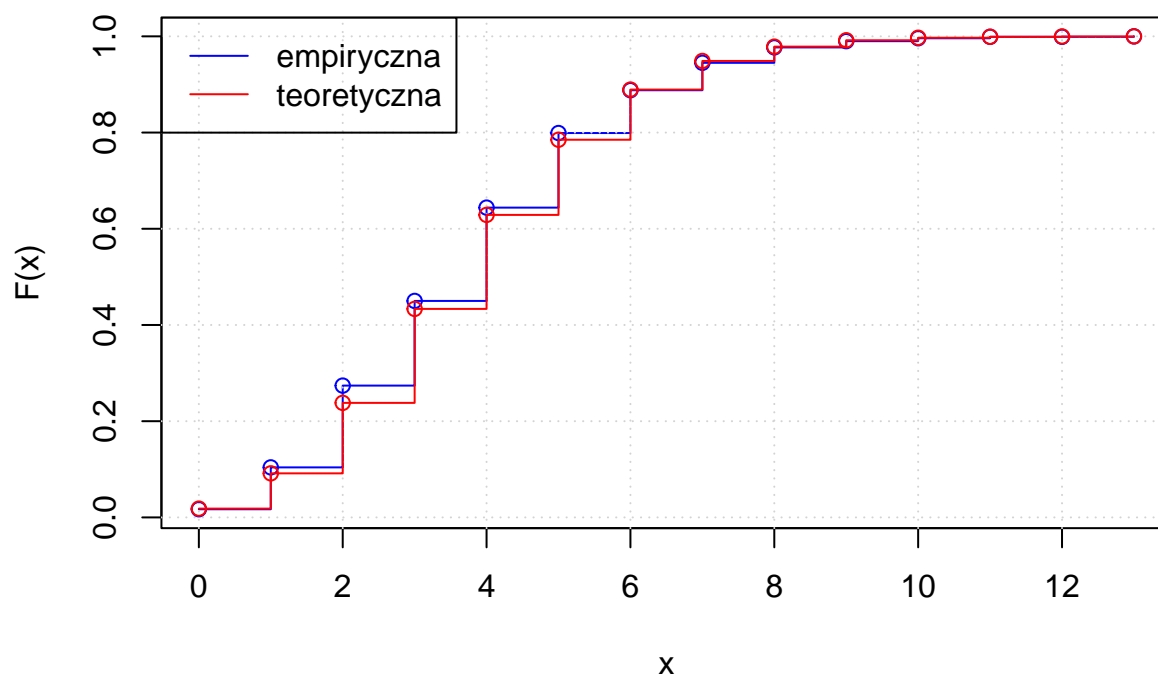
- Dystrybuanta empiryczna i teoretyczna:

```
plot(cumsum(Freq) ~ Arg, type = 's', col = 'blue',
     xlab = 'x', ylab = 'F(x)', main = paste0('Dystrybuanta dla M = ', M))
grid()
points(cumsum(Freq) ~ Arg, col = 'blue')

lines(ppois(Arg, lambda = 4) ~ Arg, type = 's', col = 'red',
     xlab = 'x', ylab = 'F(x)')
points(ppois(Arg, lambda = 4) ~ Arg, col = 'red')

legend('topleft', c('empiryczna', 'teoretyczna'),
     col = c('blue', 'red'), lwd = 1)
```

## Dystrybuanta dla $M = 1000$



- Dystrybuanta empiryczna wyznaczona za pomocą `ecdf()` z dystrybuantą teoretyczną:

```
plot(ecdf(proba), col = 'blue', xlab = 'x', ylab = 'F(x)', main = 'Dystrybuanta empiryczna (ecdf)')
lines(ppois(Arg, lambda = 4) ~ Arg, type = 's', col = 'red',
      xlab = 'x', ylab = 'F(x)')
points(ppois(Arg, lambda = 4) ~ Arg, col = 'red')

grid()
legend('topleft', c('empiryczna', 'teoretyczna'),
      col = c('blue', 'red'), lwd = 1)
```

### Dystrybuanta empiryczna (ecdf)

