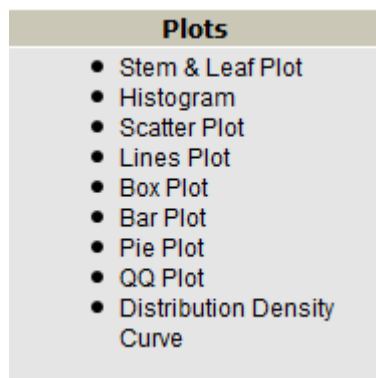


## Materiał dotyczy generowania różnego typu wykresów w środowisku R.

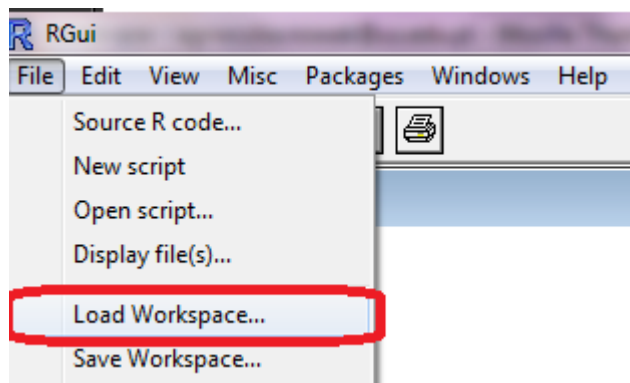
Pamiętajmy, że niektóre typy wykresów są dedykowane do pewnych typów danych.



Na potrzeby ćwiczeń początkowych załadujemy sobie zbiór danych „gnp”

<http://info.smu.edu.sg/rsite/datahandling/r/gnp.zip>

i po rozpakowaniu zapiszemy je w dowolnym miejscu na dysku. Następnie załadujemy zbiór do środowiska R.



```
> load("C:\\Users\\ANB\\Desktop\\gnp.rdata")

> print(gnp)
 [1] 516.1 514.5 517.7 513.0 517.4 527.9 538.5 551.5 564.4 572.2 579.2 582.8
 [13] 592.1 600.3 613.1 622.1 636.9 645.6 660.6 660.6
> |
```

### Wykres łodygowo-liściowy

Diagram łodygowo-listkowy (ang. stemplot lub stem-and-leaf diagram) jest graficznym sposobem prezentacji danych ilościowych. Podobnie jak histogram, służy on do przedstawiania kształtu rozkładu, ma jednak nad nim tę przewagę, że przedstawia wszystkie dane, które tworzą rozkład.

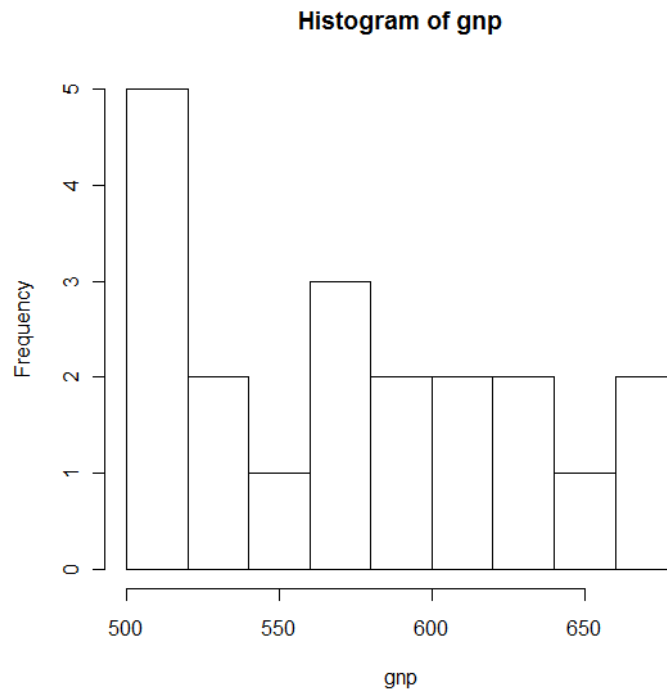
Diagram łodygowo-listkowy rysuje się w bardzo prosty sposób. Ustalamy najpierw, jakie liczby stanowić będą łodygę, zwykle opuszczając jedną lub dwie cyfry w zapisie dziesiętnym, a następnie sortujemy je rosnąco. Uzyskane liczby zapisujemy w jednej kolumnie, oddzielamy pionową kreską i dopisujemy obok obcięte końcówki – liście.

```
> stem(gnp)
```

```
The decimal point is 2 digit(s) to the right of the |
```

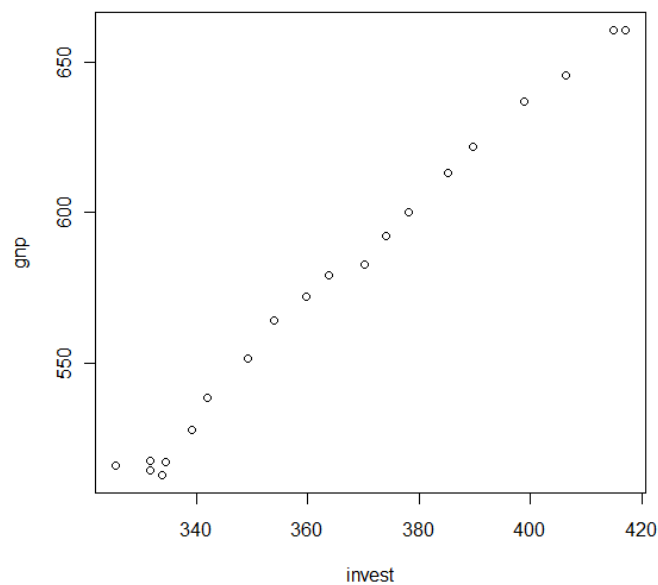
```
5 | 1122234
5 | 567889
6 | 0124
6 | 566
```

**Histogram** - wykres częstości danych. Wykonamy go za pomocą polecenia: `hist()`



## Scatter plot – wykres rozrzutu

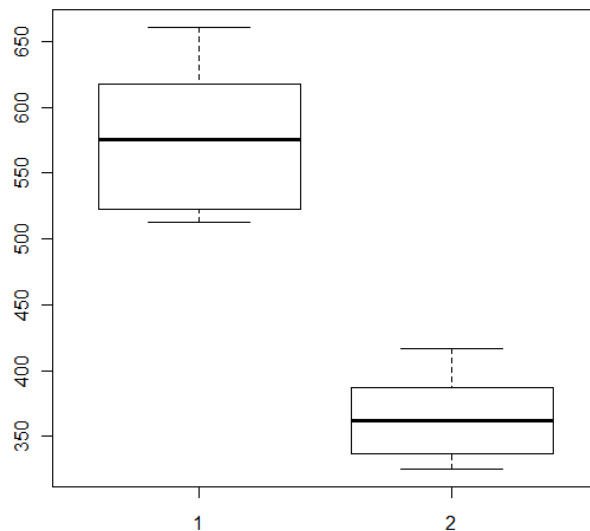
Wykres rozrzutu punktów na osiach X i Y. `plot(x-variable, y-variable)`



## Box-and-Whisker Plot – wykres pudełkowy.

`boxplot(var1, var2)` na wykresie zmienne będą zaprezentowane w takiej kolejności jak argument funkcji `boxplot` a więc najpierw `var1` a potem `var2`.

```
> boxplot(gnp,invest)
```

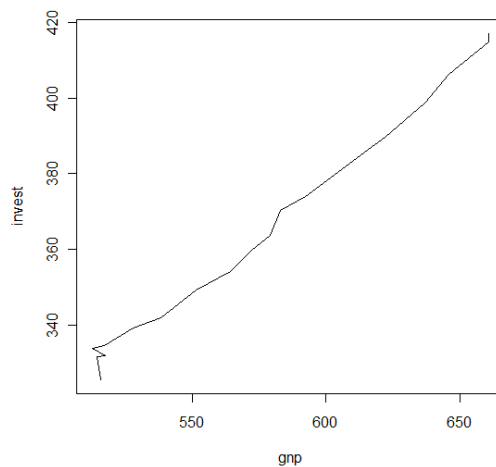


## Wykres liniowy

```
plot(gnp,invest,type="l")
```

Argument `type="l"` tworzy linię, `type="s"` będzie tworzył wykres skokowy. Poszukaj jakie jeszcze są możliwości.

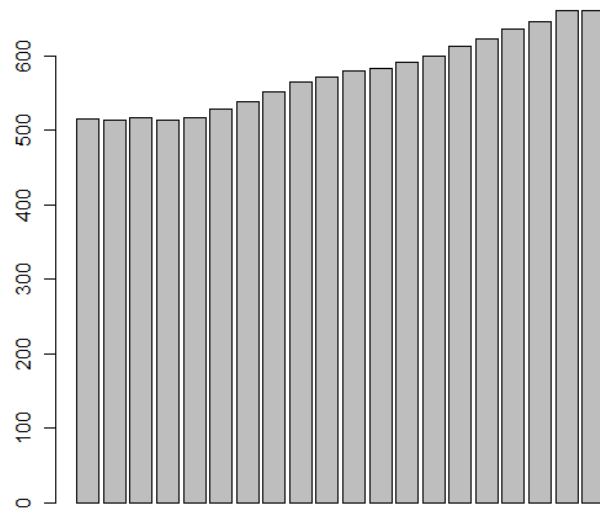
```
plot(gnp, invest, type="l")
```



## Barplot

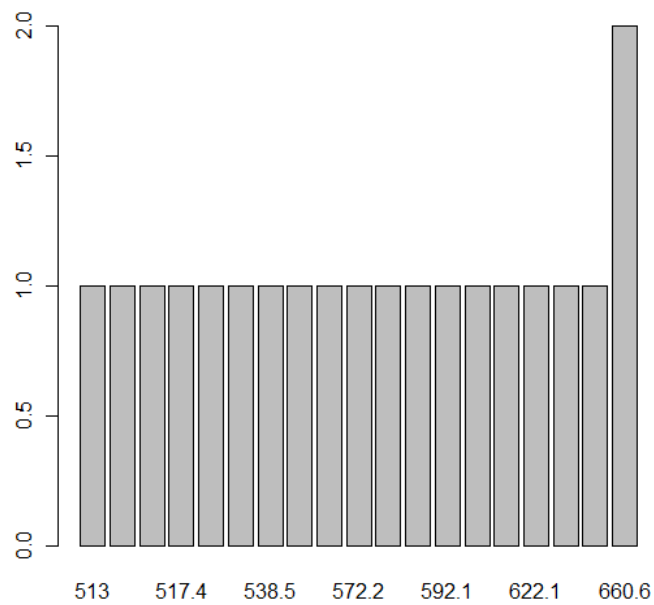
The `barplot()` - wykres kolumnowy z zaznaczeniem częstości:

```
barplot(gnp)
```



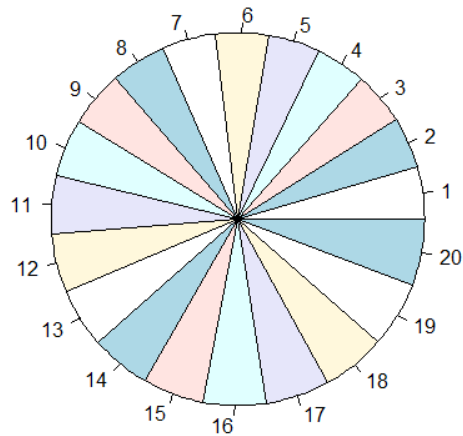
`barplot(table(var))`, `table(var)` - zwraca częstości zmiennych

`> barplot(table(gnp))`



**Wykres kołowy (Pie charts)** świetnie przedstawiają procentowy rozkład danych.

`> pie(gnp)`



Wykresy **QQ plot** oraz **distribution curve** – omówimy przy okazji regresji.

## ćwiczenia

Rozpatrzmy teraz osobno **dane jakościowe** i **ilościowe**. Najpierw zbiór w którym analizować będziemy dane jakościowe (School)

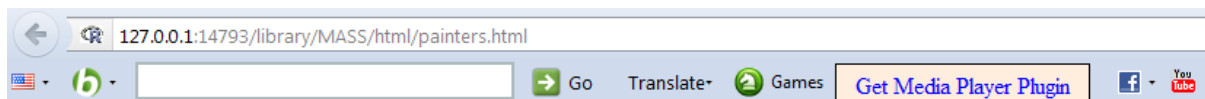
Na pierwszy rzut załaduj zbiór dotyczący malarzy.

```
> library(MASS)
> painters
```

	Composition	Drawing	Colour	Expression	School
Da Udine	10	8	16	3	A
Da Vinci	15	16	4	14	A
Del Piombo	8	13	16	7	A
Del Sarto	12	16	9	8	A
Fr. Penni	0	15	8	0	A
Guilio Romano	15	16	4	14	A
Michelangelo	8	17	4	8	A
Perino del Vaga	15	16	7	6	A
Perugino	4	12	10	4	A
Raphael	17	18	12	18	A
F. Zucarro	10	13	8	8	B
Fr. Salviata	13	15	8	8	B
Parmigiano	10	15	6	6	B
Primaticcio	15	14	7	10	B
T. Zucarro	13	14	10	9	B
Volterra	12	15	5	8	B
Barocci	14	15	6	10	C
Cortona	16	14	12	6	C

Możemy zaczerpnąć więcej wiedzy dotyczącej zbioru.

```
> help(painters)
```



painters {MASS}

## The Painter's Data of de Piles

### Description

The subjective assessment, on a 0 to 20 integer scale, of 54 classical painters. The painters were assessed on four characteristics: composition, Eighteenth century art critic, de Piles.

### Usage

painters

### Format

The row names of the data frame are the painters. The components are:

#### Composition

Composition score.

#### Drawing

Drawing score.

#### Colour

Colour score.

Ostatnia kolumna odpowiada za kod szkoły do której malarz należał. Szkoły są zakodowane jako A, B, ..., etc, - jest to cecha jakościowa.

Możemy sprawdzić jaki jest rozkład szkół

```
> painters$School
```

```
> painters$School
[1] A A A A A A A A A A B B B B B B C C C C C D D D D D D D D D D E E E E E E F F F F G G G G G G H H H H
Levels: A B C D E F G H
```

Jeśli chcemy sprawdzić rozkład częstości malarzy w poszczególnych szkołach wystarczy następujący kod:

```
> school = painters$School
> school.freq = table(school)
> school.freq
school
 A  B  C  D  E  F  G  H
10  6  6 10  7  4  7  4
> |
```

Prezentacja w postaci kolumnowej za pomocą funkcji **cbind**

```
> cbind(school.freq)
  school.freq
A           10
B              6
C              6
D           10
E              7
F              4
G              7
H              4
> |
```

ćwiczenia do wykonania:

1. Znajdź rozkład częstości dla “composition scores” w zbiorze “painters”.

Jeśli chcemy zbadać częstość względną danych jakościowych wystarczy zastosować wzór:

$$Relative\ Frequency = \frac{Frequency}{Sample\ Size}$$

```
> school.relfreq = school.freq / nrow(painters)
> school.relfreq
school
      A      B      C      D      E      F      G      H
0.18518519 0.11111111 0.11111111 0.18518519 0.12962963 0.07407407 0.12962963 0.07407407
> |
```

Możemy drukować z mniejszą liczbą cyfr i uczynić go bardziej czytelnym przez ustawienie opcji cyfr.

```
> old = options(digits=1)
> school.relfreq
school
      A      B      C      D      E      F      G      H
0.19 0.11 0.11 0.19 0.13 0.07 0.13 0.07
```

```
> cbind(school.relfreq)
school.relfreq
A      0.18518519
B      0.11111111
C      0.11111111
D      0.18518519
E      0.12962963
F      0.07407407
G      0.12962963
H      0.07407407
> old = options(digits=1)
> cbind(school.relfreq)
school.relfreq
A      0.19
B      0.11
C      0.11
D      0.19
E      0.13
F      0.07
G      0.13
H      0.07
> |
```

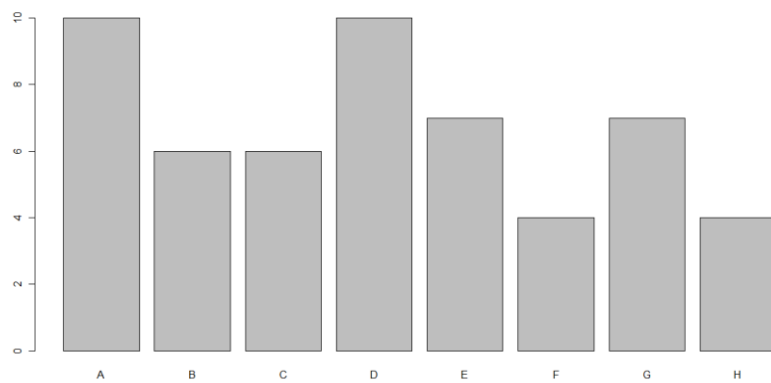
```
> options(old) # odtwarza poprzednią wersję zapisu danych
```

### ćwiczenie

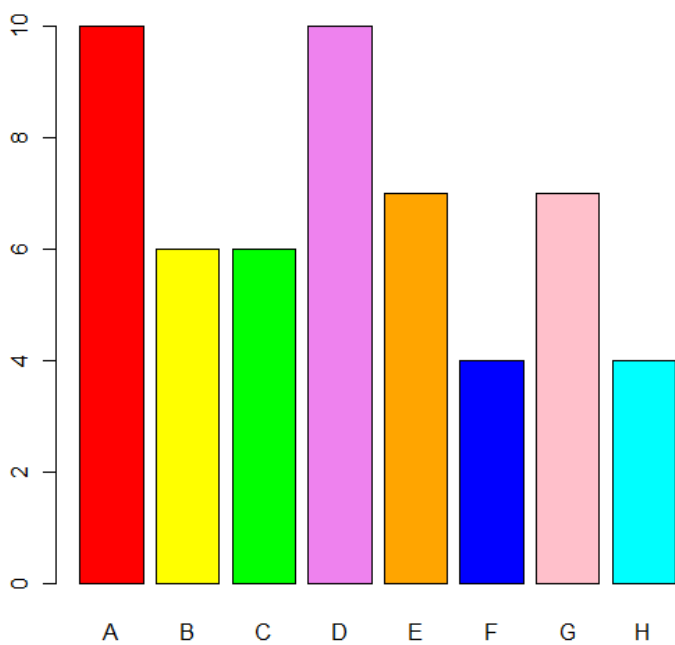
Znajdź częstość względną dla cechy „composition stores” w zbiorze painters.

### Bar plot

```
> barplot(school.freq)
> |
```



```
> colors = c("red", "yellow", "green", "violet", "orange", "blue", "pink", "cyan")
> barplot(school.freq, col=colors)
> |
```

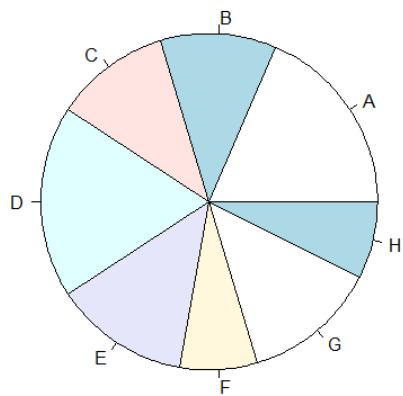


### ćwiczenie

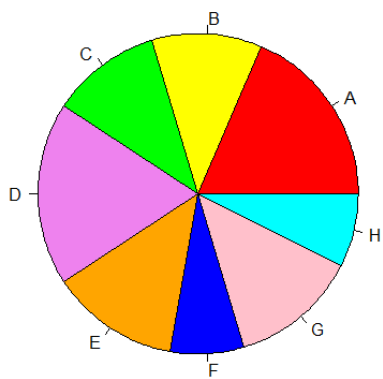
Narysuj wykres słupkowy dla cechy “composition scores” w zbiorze painters.



## Pie chart



```
> colors = c("red", "yellow", "green", "violet", "orange", "blue", "pink", "cyan")
> pie(school.freq, col=colors)
> |
```



## Ćwiczenie

Narysuj wykres typu “pie chart” dla cechy “composition scores” w zbiorze painters

## Zbiór z danymi ilościowymi

```
> head(faithful)
  eruptions waiting
1     3.600      79
2     1.800      54
3     3.333      74
4     2.283      62
5     4.533      85
6     2.883      55

> duration = faithful$eruption
> range(duration)
[1] 1.6 5.1

> breaks= seq(1.5,5.5,by=0.5)
> breaks
[1] 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5
> duration.cut = cut (duration, breaks, right = FALSE)
> duration.freq = table (duration.cut)
> duration.freq
duration.cut
[1.5,2) [2,2.5) [2.5,3) [3,3.5) [3.5,4) [4,4.5) [4.5,5) [5,5.5)
      51      41       5       7      30      73      61       4
```

```
> duration.freq = table(duration.cut)
```

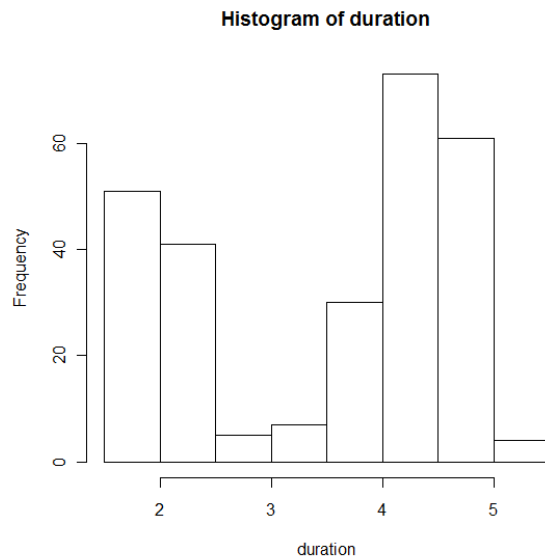
```
> cbind(duration.freq)
      duration.freq
[1.5,2)          51
[2,2.5)          41
[2.5,3)           5
[3,3.5)           7
[3.5,4)          30
[4,4.5)          73
[4.5,5)          61
[5,5.5)           4
```

### ćwiczenie

Znajdź takie same wyliczenia dla cechy “eruption waiting” w zbiorze faithful.

### Histogramy

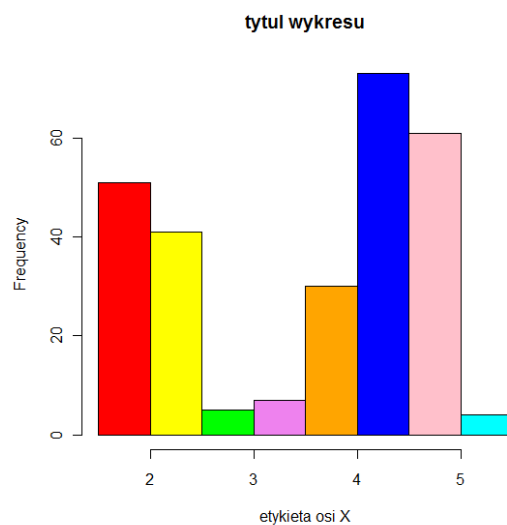
```
> duration = faithful$eruptions
> hist(duration, right=FALSE)
> |
```



ma

```
> colors = c("red", "yellow", "green", "violet", "orange", "blue", "pink", "cyan")
```

```
> hist(duration, right=FALSE, col = colors, main = "tytuł wykresu", xlab = "etykieta osi X")
```



## Ćwiczenie

Wygeneruj histogram dla cechy "eruption waiting" w zbiorze faithful.

Gdy chcemy wyznaczyć względną częstość:

```
> duration = faithful$eruptions
> breaks = seq(1.5, 5.5, by=0.5)
> duration.cut = cut(duration, breaks, right=FALSE)
> duration.freq = table(duration.cut)
> duration.relfreq = duration.freq / nrow(faithful)
> duration.relfreq
duration.cut
[1.5,2) [2,2.5) [2.5,3) [3,3.5) [3.5,4) [4,4.5) [4.5,5) [5,5.5)
  0.19   0.15   0.02   0.03   0.11   0.27   0.22   0.01
> |
```

```
> cbind(duration.freq, duration.relfreq)
      duration.freq duration.relfreq
[1.5,2)          51           0.19
[2,2.5)          41           0.15
[2.5,3)           5           0.02
[3,3.5)           7           0.03
[3.5,4)          30           0.11
[4,4.5)          73           0.27
[4.5,5)          61           0.22
[5,5.5)           4           0.01
> |
```

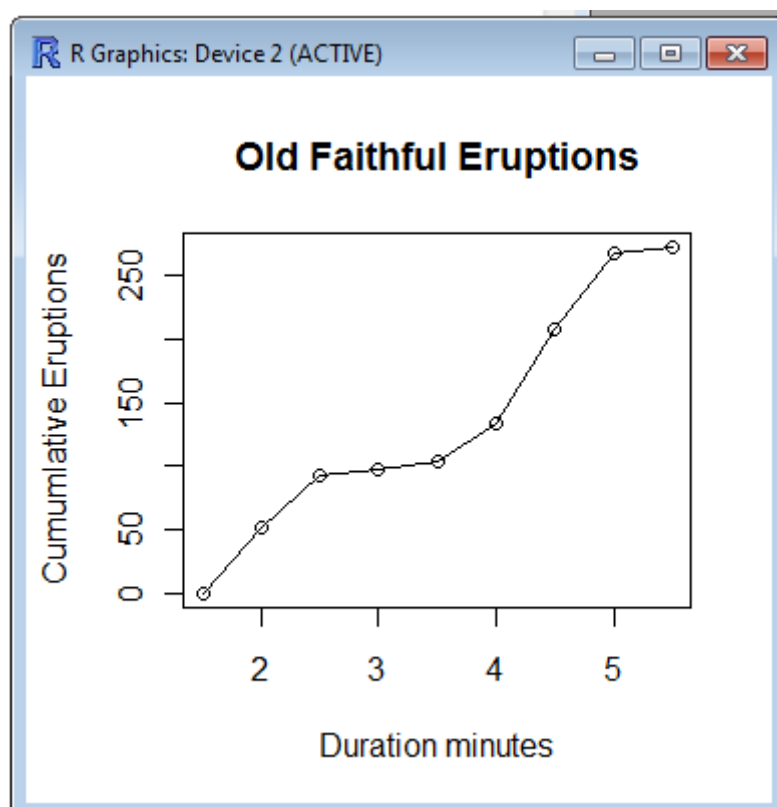
## Rozkład częstości (skumulowany)

```
> duration.cumfreq = cumsum(duration.freq)
> duration.cumfreq
[1.5,2) [2,2.5) [2.5,3) [3,3.5) [3.5,4) [4,4.5) [4.5,5) [5,5.5)
      51      92      97      104      134      207      268      272
> |
```

## ćwiczenie

Wyznacz skumulowany rozkład częstości dla cechy “eruption Whiting” w zbiorze faithful.

```
> cumfreq0 = c(0, cumsum(duration.freq))
> plot(breaks, cumfreq0, main="Old Faithful Eruptions", xlab="Duration minutes", ylab="Cumulative Eruptions")
> lines(breaks, cumfreq0)
> |
```



## ćwiczenie

Wykreśl taki sam wykres tyle, że dla cechy “eruption Waiting” w zbiorze faithful.

### Względny rozkład częstości (skumulowany)

$$\text{Cumulative Relative Frequency} = \frac{\text{Cumulative Frequency}}{\text{Sample Size}}$$

```
> duration.cumrelfreq = duration.cumfreq / nrow(faithful)
> duration.cumrelfreq
[1.5,2) [2,2.5) [2.5,3) [3,3.5) [3.5,4) [4,4.5) [4.5,5) [5,5.5)
      0.2      0.3      0.4      0.4      0.5      0.8      1.0      1.0
. |
```

## Ćwiczenie

Wykonaj te same działania dla cechy “eruption waiting” w zbiorze faithful.

## Wykres łodygowo-liściowy

Diagram łodygowo-listkowy (ang. stemplot lub stem-and-leaf diagram) jest graficznym sposobem prezentacji danych ilościowych. Podobnie jak histogram, służy on do przedstawiania kształtu rozkładu, ma jednak nad nim tę przewagę, że przedstawia wszystkie dane, które tworzą rozkład.

Diagram łodygowo-listkowy rysuje się w bardzo prosty sposób. Ustalamy najpierw, jakie liczby stanowią będą łodygę, zwykle opuszczając jedną lub dwie cyfry w zapisie dziesiętnym, a następnie sortujemy je rosnąco. Uzyskane liczby zapisujemy w jednej kolumnie, oddzielamy pionową kreską i dopisujemy obok obcięte końcówki – liście.

```
> duration = faithful$eruptions
> stem(duration)

The decimal point is 1 digit(s) to the left of the |

16 | 070355555588
18 | 0000222333333355777777777888822335777888
20 | 00002223378800035778
22 | 0002335578023578
24 | 00228
26 | 23
28 | 080
30 | 7
32 | 2337
34 | 250077
36 | 0000823577
38 | 2333335582225577
40 | 0000003357788888002233555577778
42 | 03335555778800233333555577778
44 | 02222335557780000000023333357778888
46 | 00002333577000000023578
48 | 00000022335800333
50 | 0370

> |
```

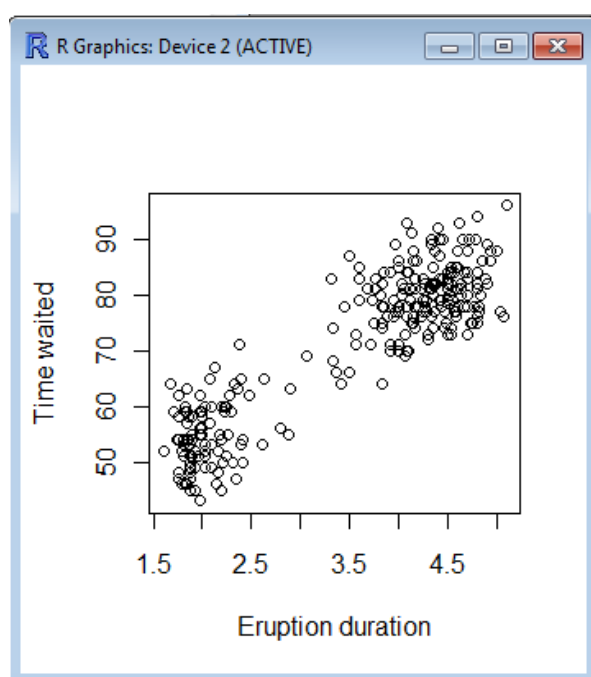
## Ćwiczenie

Wykreśl taki sam wykres dla cechy "eruption waiting" w zbiorze faithful.

### Wykresy rozrzutu (ang. Scatter Plot)

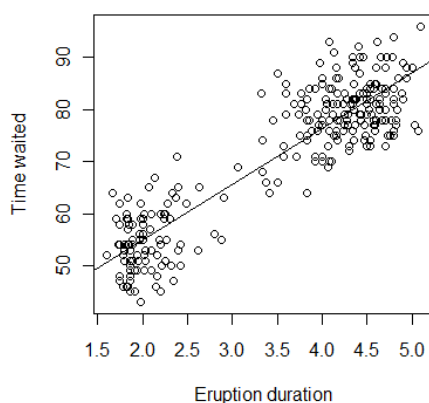
```
> duration = faithful$eruptions
> waiting = faithful$waiting
> head(cbind(duration, waiting))
      duration waiting
[1,]         4      79
[2,]         2      54
[3,]         3      74
[4,]         2      62
[5,]         5      85
[6,]         3      55
> |

> plot(duration, waiting,xlab="Eruption duration",ylab="Time waited")
> |
```



Dodanie linii regresji do wykresu

```
> 
> plot(duration, waiting,xlab="Eruption duration",ylab="Time waited")
> abline(lm(waiting ~ duration))
> |
```



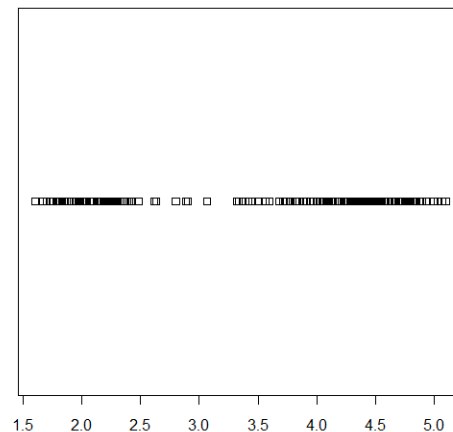
## Wykresy typu paskowego

Wykres paskowy (ang. strip chart) można otrzymać wydając polecenie **stripchart**

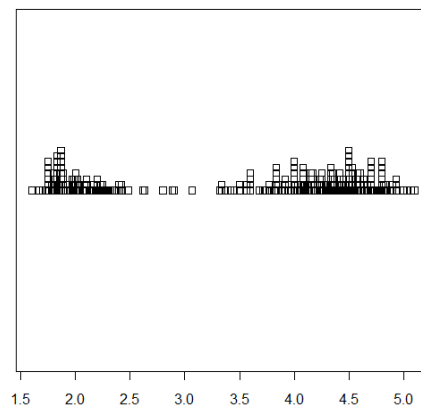
Wydając to polecenie z opcjami `pch=16` (obserwacje ze zbioru danych będą zaznaczone przy użyciu kropek), `method="stack"` (obserwacje powtarzające się będą "ustawiane" jedna nad drugą), `cex=2` (kropki zostaną odpowiednio powiększone):  
`stripchart(y,method="stack",pch=16,cex=2)`

**> stripchart(duration)**

Efekt jest następujący:



`stripchart(duration,method="stack")`

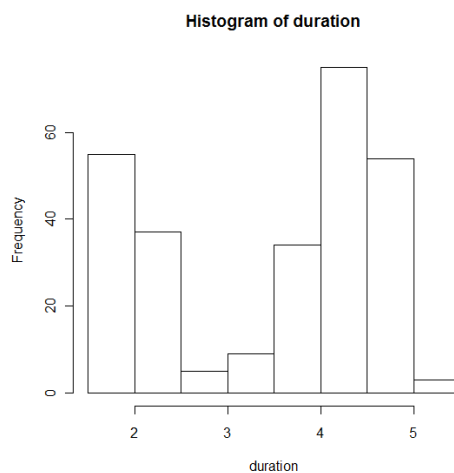


Proszę sprawdzić opcję `stripchart(duration,method="jitter")`

- `stripchart(duration,vertical=TRUE)`
- `stripchart(w1$vals,vertical=TRUE,method="jitter")`
- `stripchart(w1$vals,method="stack", main='Leaf BioMass in High CO2 Environment',  
xlab='BioMass of Leaves')`

## HISTOGRAMY

```
hist(duration)
```



Jak widać przedziały zostały automatycznie określone, ale zawsze można je regulować.

```
> hist(duration)
> hist(duration,breaks = 2)
> hist(duration,breaks = 4)
> hist(duration,breaks = 8)
> hist(duration,breaks = 12)
> |
```

Można także decydować o rozpiętości danych na osi x:

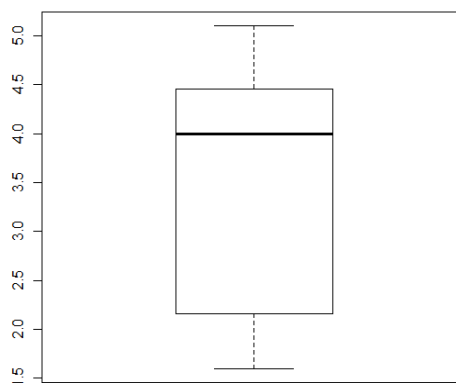
```
hist(duration,breaks=12,xlim=c(0,10))
```

Proszę sprawdzić także następujące przykłady:

```
> hist(duration,breaks=12,xlim=c(-1,2))
> hist(duration,breaks=12,xlim=c(0,2))
> hist(duration,breaks=12,xlim=c(1,1.3))
> hist(duration,breaks=12,xlim=c(0.9,1.3))
```

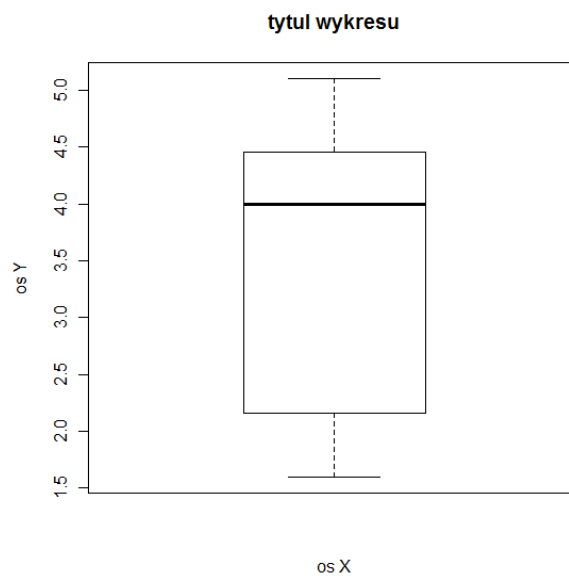
## Wykresy typu Boxplots

```
boxplot(w1)
```

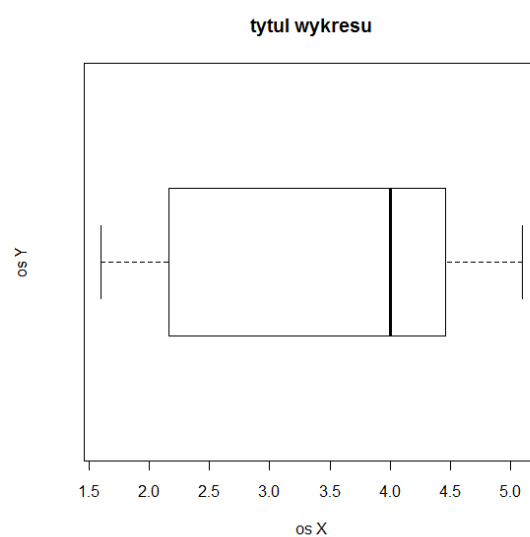




```
> boxplot(duration,main='tytuł wykresu',xlab='os X', ylab = 'os Y')
```



```
> boxplot(w1,main='tytuł wykresu',xlab='os X', ylab = 'os Y',horizontal=TRUE)
```



A sprawdź co będzie efektem następującego wywołania obu linii (podstaw za "w1\$vals" -> „duration”)

```
> hist(w1$vals,main='tytuł wykresu',xlab='os X',ylim=c(0,16))  
> boxplot(w1$vals,horizontal=TRUE,at=15.5,add=TRUE,axes=FALSE)
```