Relative Frequency distribution

Relative Frequency distribution

It is a summary of the frequency proportion in a collection of non-overlapping categories.

Weights (in kg)	Frequency	Relative Frequency
50	10	$\frac{10}{50}=0.2$
60	12	$\frac{12}{50}=0.24$
70	5	$\frac{5}{50}=0.1$
55	13	$\frac{13}{50}=0.26$
40	10	$\frac{10}{50}=0.2$
Total	50	

$$\text{Relative frequency} = \frac{\text{count of subgroup}}{\text{Total count}} \times 100$$

https://www.chegg.com/homework-help/definitions/relative-frequency-distribution-31

Relative Frequency distribution

```
relschooltable=schooltable/nrow(painters)
```

```
> relschooltable=schooltable/nrow(painters)
> relschooltable
school
         A          B          C          D          E
0.18518519 0.11111111 0.11111111 0.18518519 0.12962963
         F          G          H
0.07407407 0.12962963 0.07407407
> cbind(relschooltable)
  relschooltable
A     0.18518519
B     0.11111111
C     0.11111111
D     0.18518519
E     0.12962963
F     0.07407407
G     0.12962963
H     0.07407407
>
```

```
library(MASS)
```

Relative frequency distribution

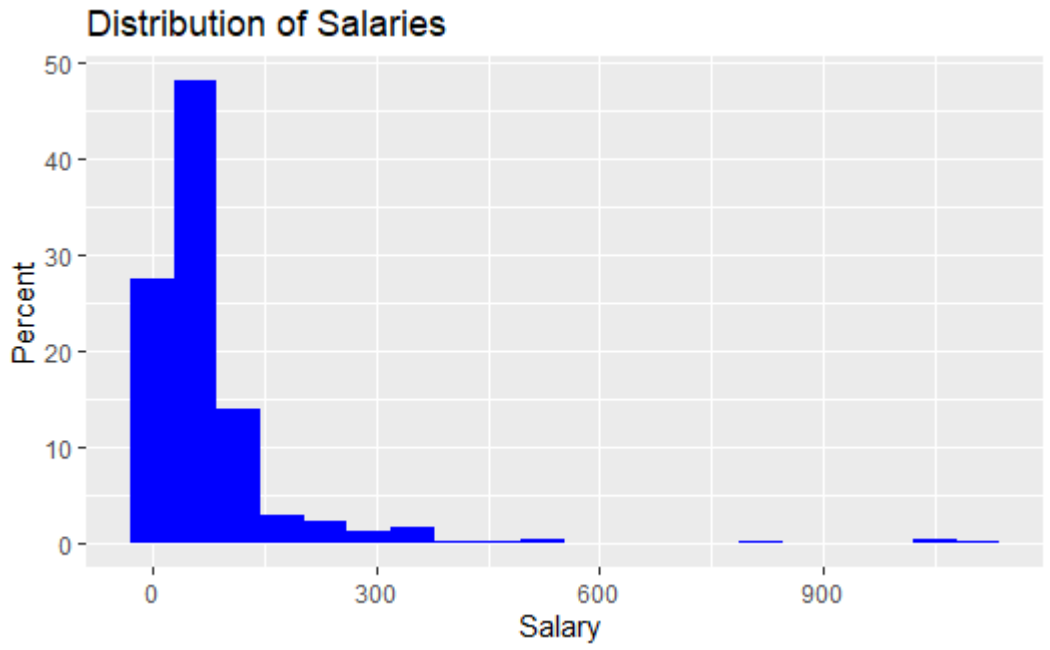
$$\text{Relative frequency} = \frac{\text{count of subgroup}}{\text{Total count}} \times 100$$

```
reltableSalary=tableSalary/nrow(mydata)*100
```

```
reltableSalary=tableSalary/sum(tableSalary)*100
```

Relative frequency distribution

```
ggplot(data=mydata) +
  geom_histogram(mapping=aes(x=salary, y=..count../sum(..count..)*100), bins=20, fill="blue") +
  ggtitle("Distribution of Salaries") +
  xlab("Salary ") +
  ylab("Percent")
```



Cumulative frequency distribution

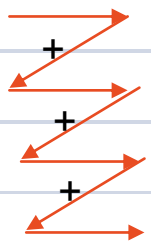
cumulative frequency distribution of a quantitative variable is a summary of data frequency below given levels.

Class Limits	Frequency	Cumulative Frequency
5-10	1	1
10-15	2	3
15-20	4	7
20-25	0	7
25-30	3	10
30-35	5	15
35-40	6	21

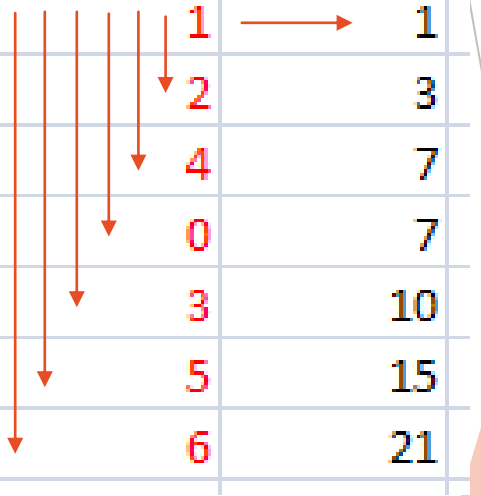
cumsum()

cumsum(tableSalary)

Class Limits	Frequency	Cumulative Frequency
5-10	1	1
10-15	2	3
15-20	4	7
20-25	0	7
25-30	3	10
30-35	5	15
35-40	6	21



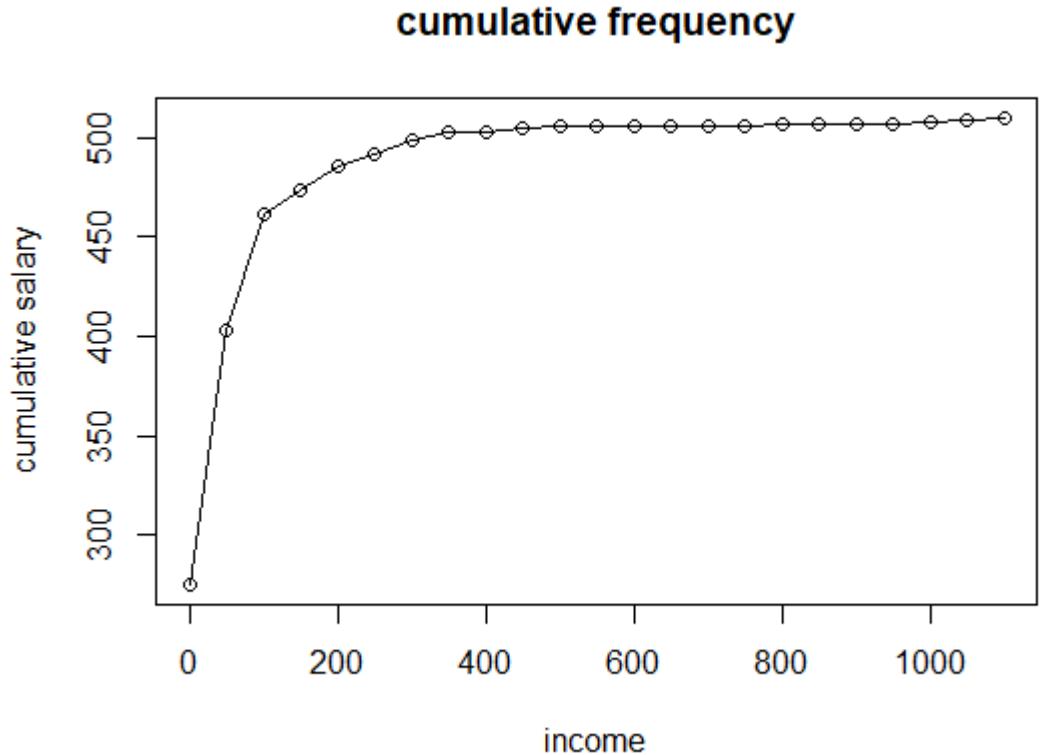
Class Limits	Frequency	Cumulative Frequency
5-10	1	1
10-15	2	3
15-20	4	7
20-25	0	7
25-30	3	10
30-35	5	15
35-40	6	21



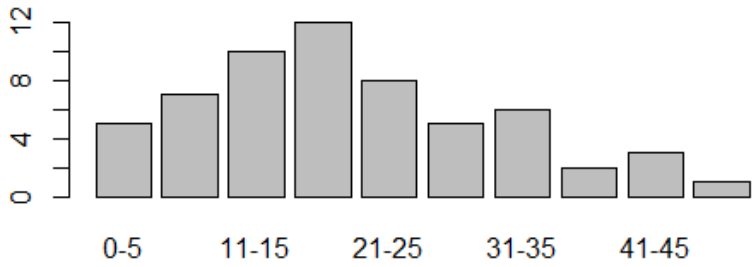
Cumulative frequency graph

cumulative frequency distribution of a quantitative variable is a summary of data frequency below given levels.

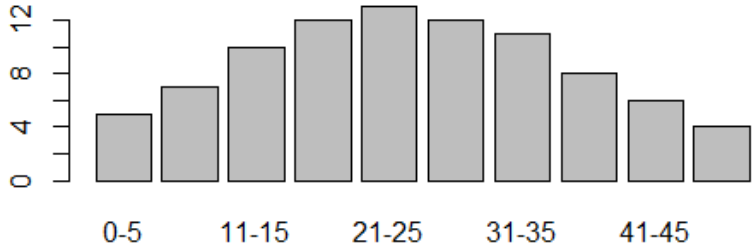
```
plot(breaks[1:23],cumtabSalary, main = "cumulative
frequency", xlab = "income", ylab = "cumulative
salary")

 lines(breaks[1:23],cumtabSalary)
```



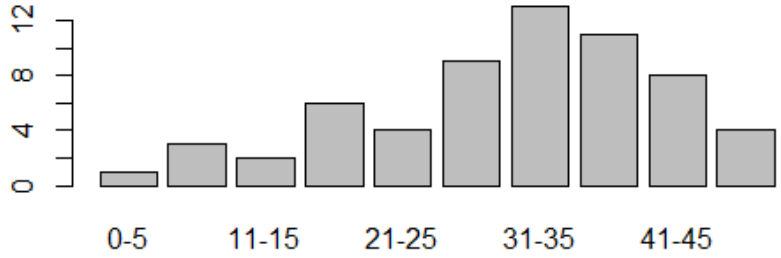
```
> cbind(tmpage)
      tmpage
0-5        5
6-10       7
11-15     10
16-20     12
21-25      8
26-30      5
31-35      6
36-40      2
41-45      3
46-50      1
```




```
> cbind(tmpage2)
      tmpage2
0-5         5
6-10        7
11-15      10
16-20      12
21-25      13
26-30      12
31-35      11
36-40       8
41-45       6
46-50       4
```




```
> cbind(tmpage3)
      tmpage3
0-5         1
6-10        3
11-15       2
16-20       6
21-25       4
26-30       9
31-35      13
36-40      11
41-45       8
46-50       4
```

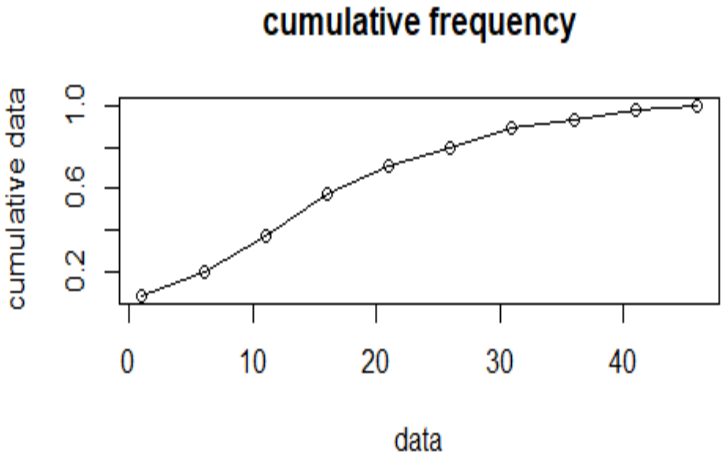


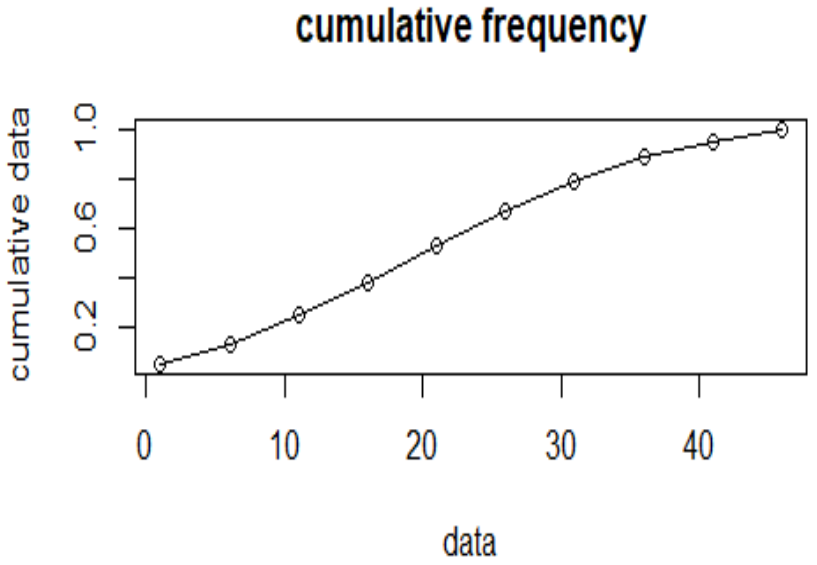
Relative frequency

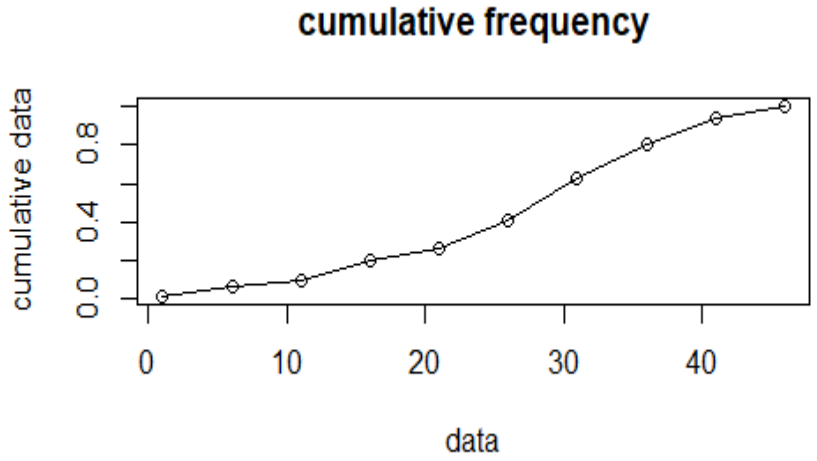
```
> cbind(relfreq1)
        relfreq1
0-5   0.08474576
6-10  0.11864407
11-15 0.16949153
16-20 0.20338983
21-25 0.13559322
26-30 0.08474576
31-35 0.10169492
36-40 0.03389831
41-45 0.05084746
46-50 0.01694915
> sum(relfreq1)
[1] 1
```

Cumulative frequency distribution

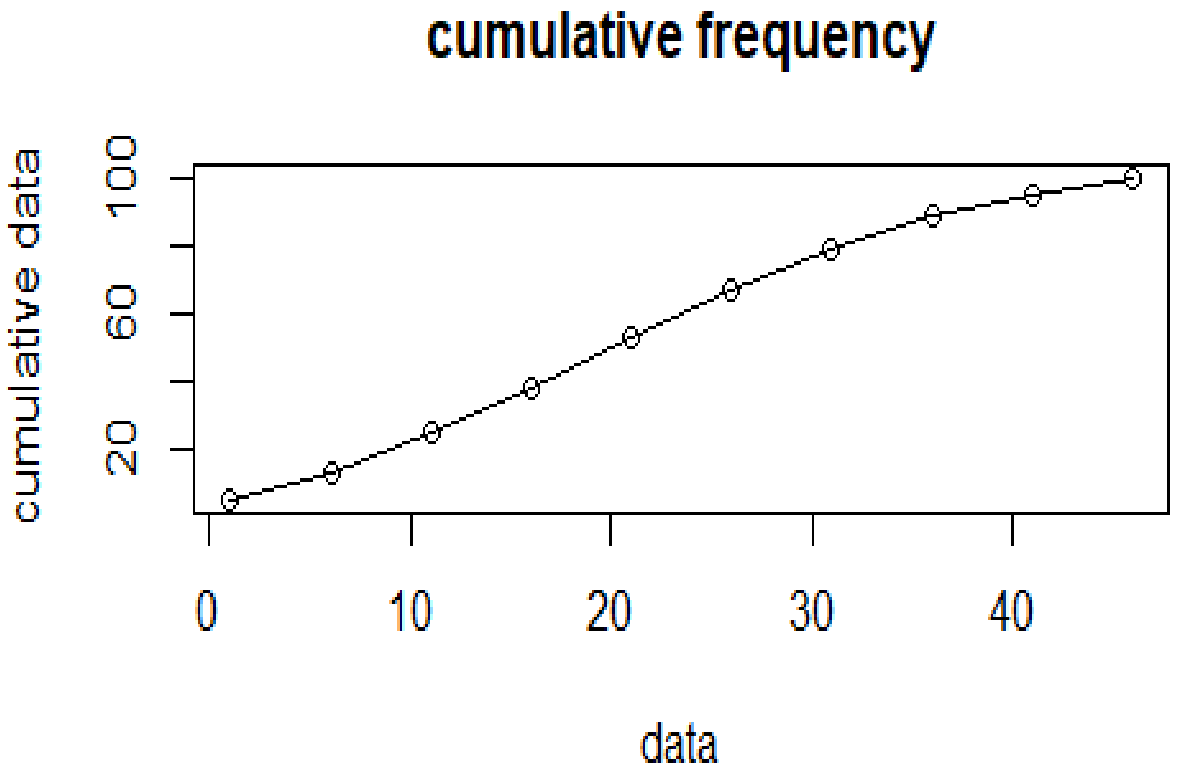
```
> cbind(cumRF1)
          cumRF1
0-5   0.08474576
6-10  0.20338983
11-15 0.37288136
16-20 0.57627119
21-25 0.71186441
26-30 0.79661017
31-35 0.89830508
36-40 0.93220339
41-45 0.98305085
46-50 1.00000000
```


cumulative frequency
cumulative data
0.2
0.6
1.0
0
10
20
30
40
data


cumulative frequency
cumulative data
0.2
0.6
1.0
0
10
20
30
40
data


cumulative frequency
cumulative data
0.0
0.4
0.8
0
10
20
30
40
data

```
> cbind(cumRF2*100)
             [,1]
0-5      5.681818
6-10    13.636364
11-15   25.000000
16-20   38.636364
21-25   53.409091
26-30   67.045455
31-35   79.545455
36-40   88.636364
41-45   95.454545
46-50  100.000000
```

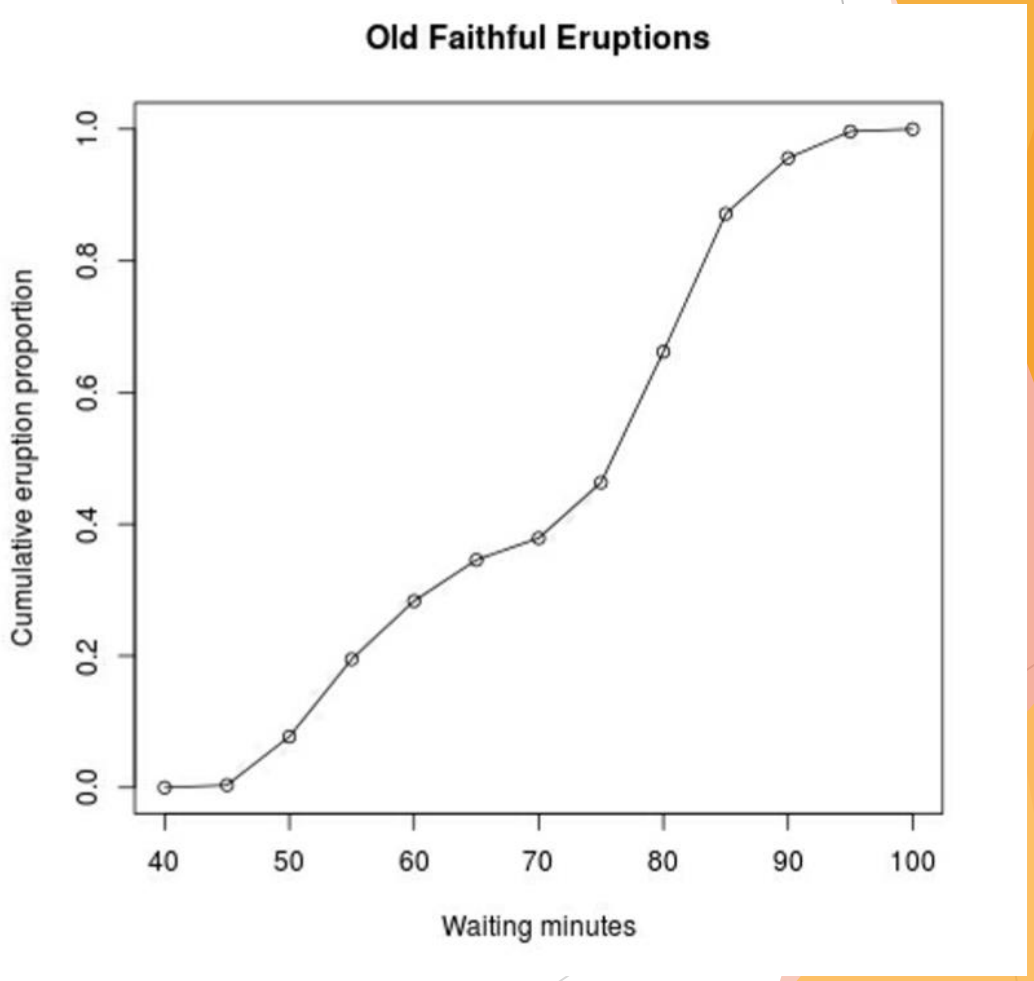


Exercise

head(faithful)

?faithful

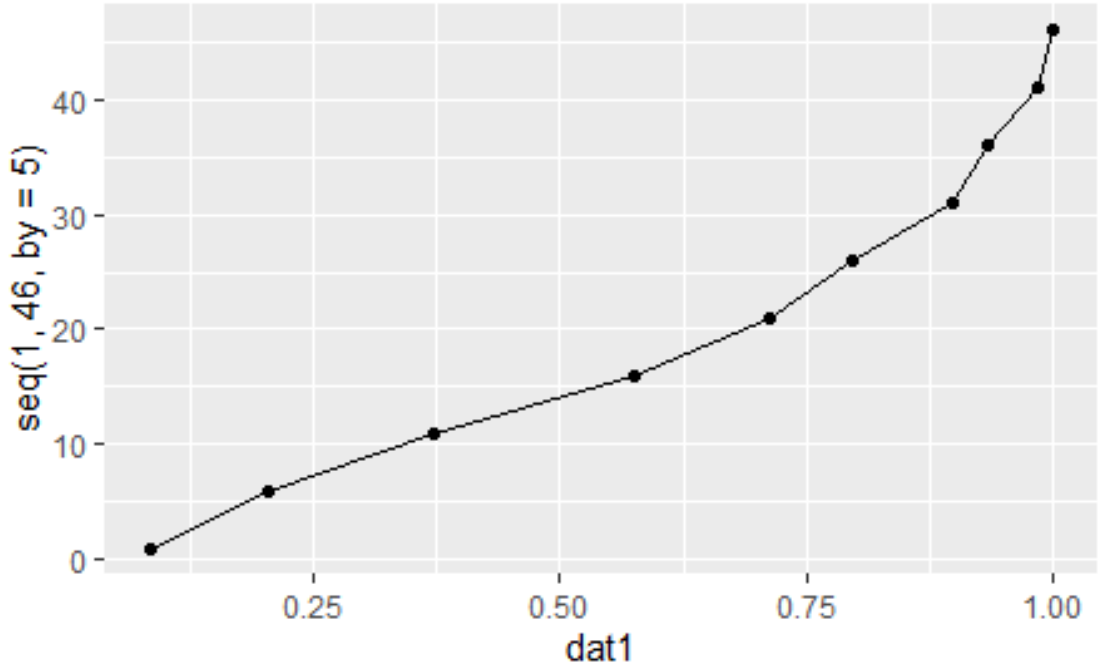
Find the cumulative relative frequency graph of the eruption durations in faithful.



Line graph with ggplot2

```
mydatFrame2=data.frame(dat1=cumRF1,dat2=cumRF2,dat3=cumRF3)
```

```
ggplot(data=mydatFrame2, aes(x=dat1, y=seq(1,46, by=5))) +
  geom_line()+
  geom_point()
```

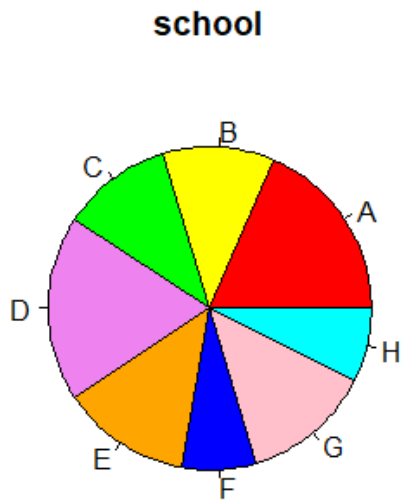


Pie chart

consists of pizza wedges that shows the frequency distribution graphically.

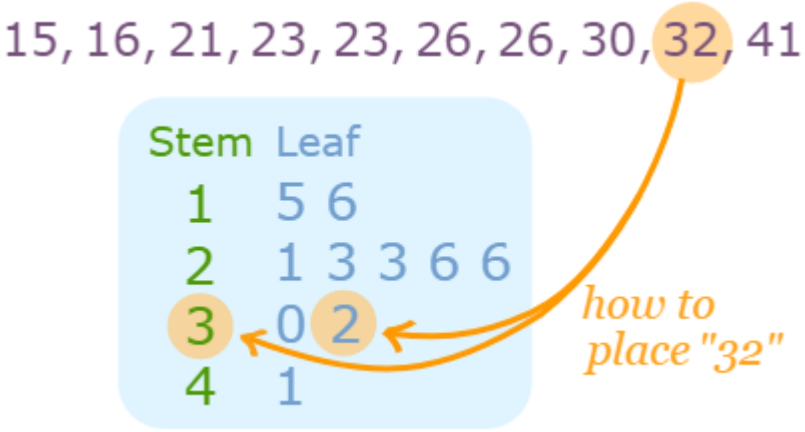
```
pie(schooltable)
```

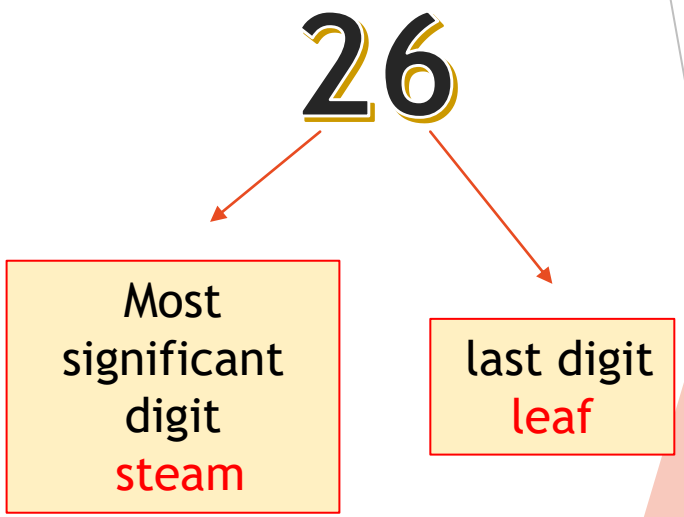
```
pie(schooltable, main= "school",col=colors)
```



Stem and leaf plot

A stem-and-leaf plot of a quantitative variable is a textual graph that classifies data items according to their most significant numeric digits. In addition, we often merge each alternating row with its next row in order to simplify the graph for readability.





Stem and leaf plot

stem(faithful$eruptions)

```
[1] 3.600 1.800 3.333 2.283 4.533 2.883 4.700 3.600 1.950 4.350 1.833 3.917 4.200
 [14] 1.750 4.700 2.167 1.750 4.800 1.600 4.250 1.800 1.750 3.450 3.067 4.533 3.600
 [27] 1.967 4.083 3.850 4.433 4.300 4.467 3.367 4.033 3.833 2.017 1.867 4.833 1.833
 [40] 4.783 4.350 1.883 4.567 1.750 4.533 3.317 3.833 2.100 4.633 2.000 4.800 4.716
 [53] 1.833 4.833 1.733 4.883 3.717 1.667 4.567 4.317 2.233 4.500 1.750 4.800 1.817
 [66] 4.400 4.167 4.700 2.067 4.700 4.033 1.967 4.500 4.000 1.983 5.067 2.017 4.567
 [79] 3.883 3.600 4.133 4.333 4.100 2.633 4.067 4.933 3.950 4.517 2.167 4.000 2.200
 [92] 4.333 1.867 4.817 1.833 4.300 4.667 3.750 1.867 4.900 2.483 4.367 2.100 4.500
[105] 4.050 1.867 4.700 1.783 4.850 3.683 4.733 2.300 4.900 4.417 1.700 4.633 2.317
[118] 4.600 1.817 4.417 2.617 4.067 4.250 1.967 4.600 3.767 1.917 4.500 2.267 4.650
[131] 1.867 4.167 2.800 4.333 1.833 4.383 1.883 4.933 2.033 3.733 4.233 2.233 4.533
[144] 4.817 4.333 1.983 4.633 2.017 5.100 1.800 5.033 4.000 2.400 4.600 3.567 4.000
[157] 4.500 4.083 1.800 3.967 2.200 4.150 2.000 3.833 3.500 4.583 2.367 5.000 1.933
[170] 4.617 1.917 2.083 4.583 3.333 4.167 4.333 4.500 2.417 4.000 4.167 1.883 4.583
[183] 4.250 3.767 2.033 4.433 4.083 1.833 4.417 2.183 4.800 1.833 4.800 4.100 3.966
[196] 4.233 3.500 4.366 2.250 4.667 2.100 4.350 4.133 1.867 4.600 1.783 4.367 3.850
[209] 1.933 4.500 2.383 4.700 1.867 3.833 3.417 4.233 2.400 4.800 2.000 4.150 1.867
[222] 4.267 1.750 4.483 4.000 4.117 4.083 4.267 3.917 4.550 4.083 2.417 4.183 2.217
[235] 4.450 1.883 1.850 4.283 3.950 2.333 4.150 2.350 4.933 2.900 4.583 3.833 2.083
[248] 4.367 2.133 4.350 2.200 4.450 3.567 4.500 4.150 3.817 3.917 4.450 2.000 4.283
[261] 4.767 4.533 1.850 4.250 1.983 2.250 4.750 4.117 2.150 4.417 1.817 4.467
```

Key 30| 7 is 3.07

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

```
16 | 070355555588
18 | 0000222333333355777777778888822335777888
20 | 00002223378800035778
22 | 000233578023578
24 | 00228
26 | 23
28 | 080
30 | 7
32 | 2337
34 | 250077
36 | 0000823577
38 | 2333335582225577
40 | 000000335778888800223355557777 8
42 | 033355557788002333335555777 78
44 | 022223355577800000000233333577788 88
46 | 000023335770000002357 8
48 | 0000002233580033 3
50 | 0370
```