STK 573
Metode Grafik untuk Analisis dan Penyajian Data

Pertemuan 6
Sajian Peubah Kontinu Tunggal

Tim Dosen:
Dr. Bagus Sartono
Dr. Ir. Aji Hamim Wigena
Dr. Agus M Soleh

Introduction

- exploring distributions of a single continuous variable
- continuous variables as plots
- discrete (categorical) variables as charts

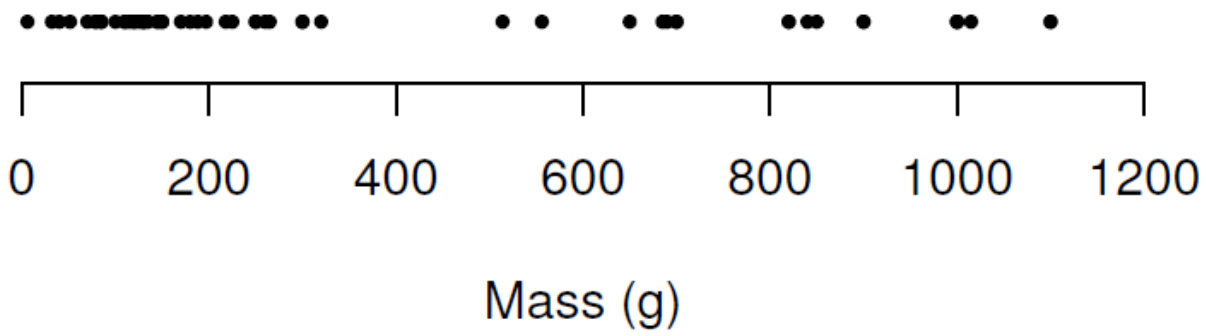
Introduction

- The Dotplot
- The Stemplot
- The Boxplot
- The EDF Plot

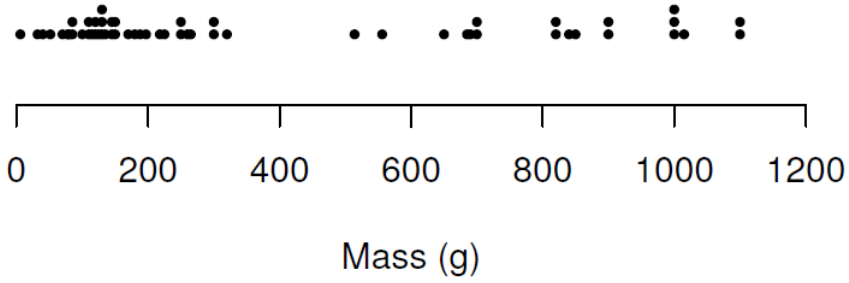
The Dotplot

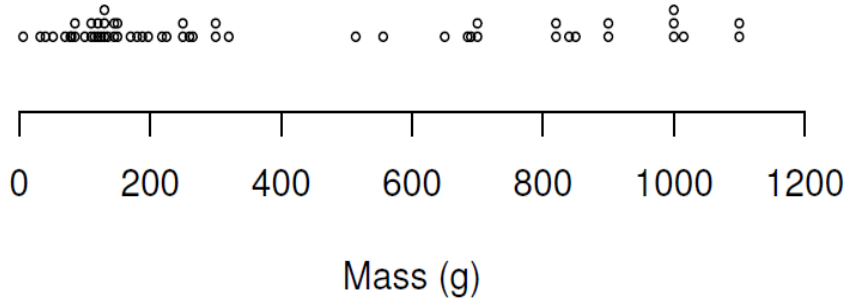
- one-dimensional scatterplots
- referred to as a stripchart
- Original:
  - Replicated observations occur as a single dot
  - If observations occur sufficiently close together the separation between dots is not apparent at a given scale and they can also appear to be a single dot
- function: stripchart()

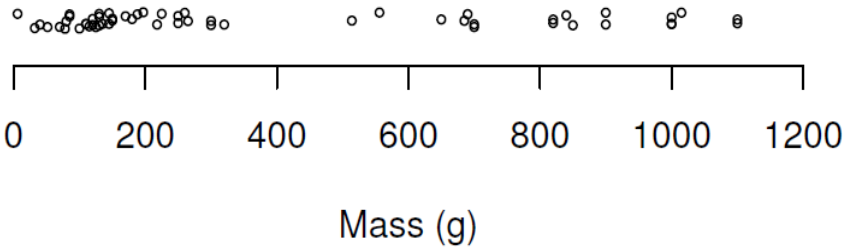
Mass (g)							
5.9	32.0	40.0	51.5	70.0	100.0	78.0	80.0
85.0	85.0	110.0	115.0	125.0	130.0	120.0	120.0
130.0	135.0	110.0	130.0	150.0	145.0	150.0	170.0
225.0	145.0	188.0	180.0	197.0	218.0	300.0	260.0
265.0	250.0	250.0	300.0	320.0	514.0	556.0	840.0
685.0	700.0	700.0	690.0	900.0	650.0	820.0	850.0
900.0	1,015.0	820.0	1,100.0	1,000.0	1,100.0	1,000.0	1,000.0

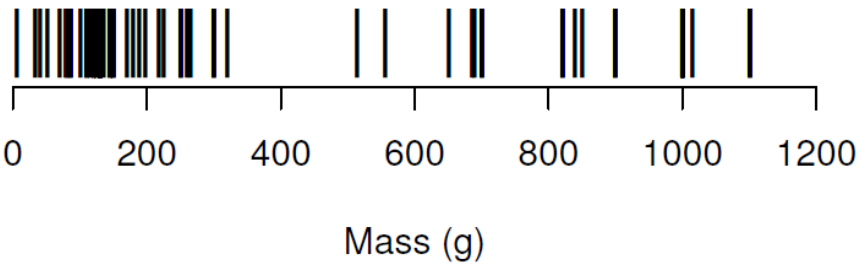


- Variations on the Dotplot:
  - stacked dotplot -> identical observations can be ameliorated
  - open circles -> eliminate smear over a range of values
  - Jittering
  - barcode plot -> dot icon replaced by a vertical line segment


0
200
400
600
800
1000
1200
Mass (g)


0
200
400
600
800
1000
1200
Mass (g)


0
200
400
600
800
1000
1200
Mass (g)


0
200
400
600
800
1000
1200
Mass (g)

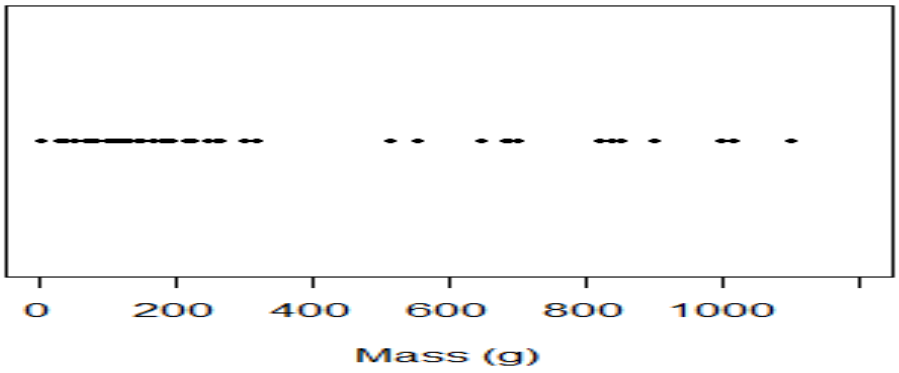
Praktek di R

Data

Mass (gram)							
5.9	32	40	51.5	70	100	78	80
85	85	110	115	125	130	120	120
130	135	110	130	150	145	150	170
225	145	188	180	197	218	300	260
265	250	250	300	320	514	556	840
685	700	700	690	900	650	820	850
900	1015	820	1100	1000	1100	1000	1000

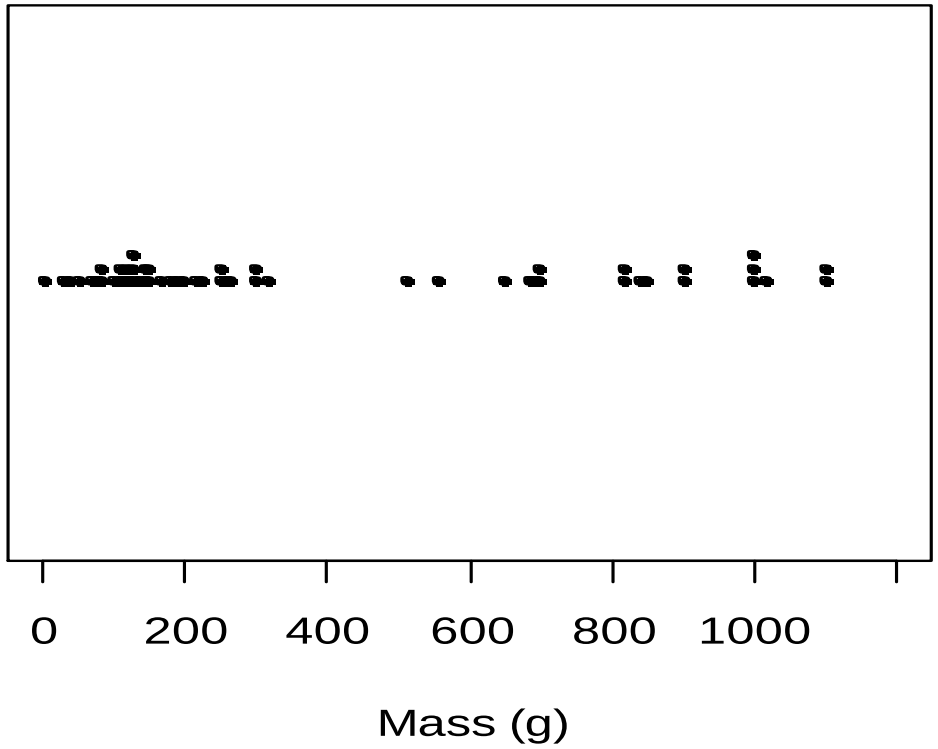
Dot Plot

```
mass<-c(5.9,32.0,40.0,51.5,70.0,100.0,78.0,80.0,85.0,85.0,
110.0,115.0,125.0,130.0,120.0,120.0,130.0,135.0,110.0,130.0,
150.0,145.0,150.0,170.0,225.0,145.0,188.0,180.0,197.0,218.0,
300.0,260.0,265.0,250.0,250.0,300.0,320.0,514.0,556.0,840.0,
685.0,700.0,700.0,690.0,900.0,650.0,820.0,850.0,900.0,1015.0,
820.0,1100.0,1000.0,1100.0,1000.0,1000.0)
#
stripchart(mass,method="overplot",xlab="Mass (g)",pch=19,cex=0.5,
xlim=c(0.,1200.))
```



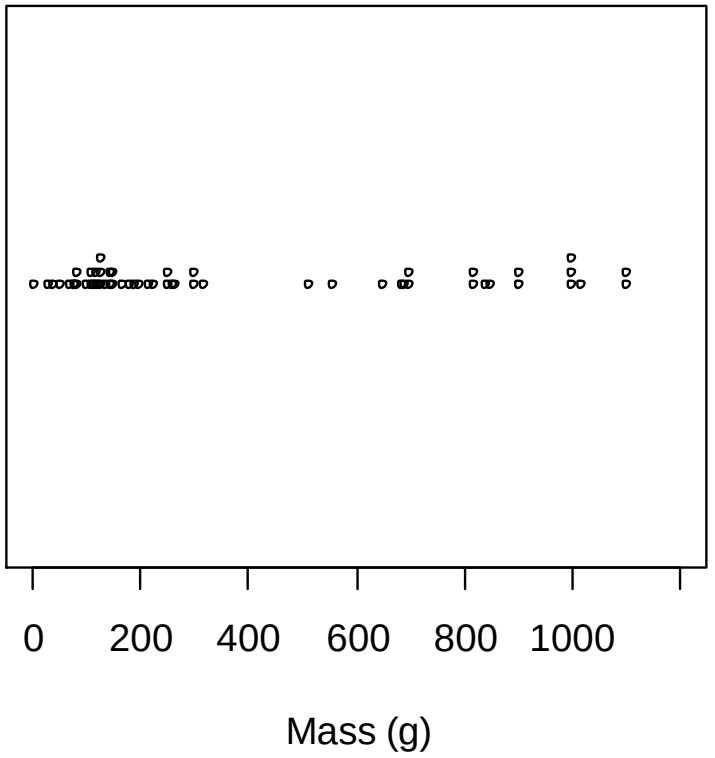
Stacked Dot Plot

```
stripchart(mass,method="stack",of
fset=0.6,xlab="Mass
(g)",pch=19,cex=0.5,
xlim=c(0.,1200.))
```



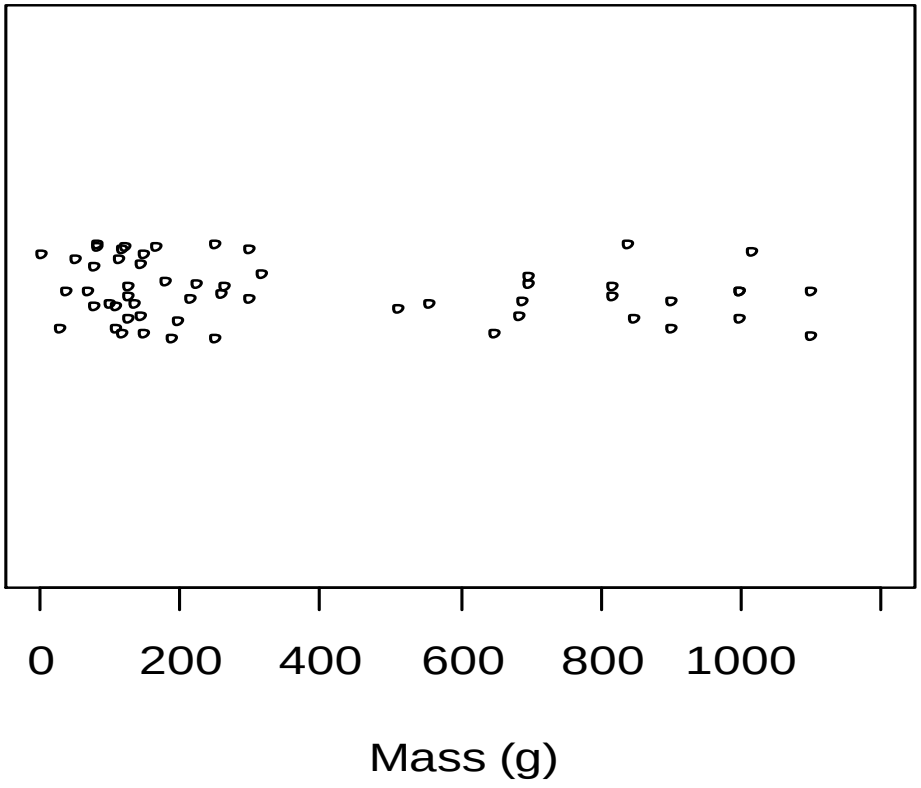
Stacked Dot Plot with Open Circle Replacing Dots

```
stripchart(mass,method="stack",of
fset=0.6,xlab="Mass
(g)",pch=21,cex=0.5,
xlim=c(0.,1200.))
```



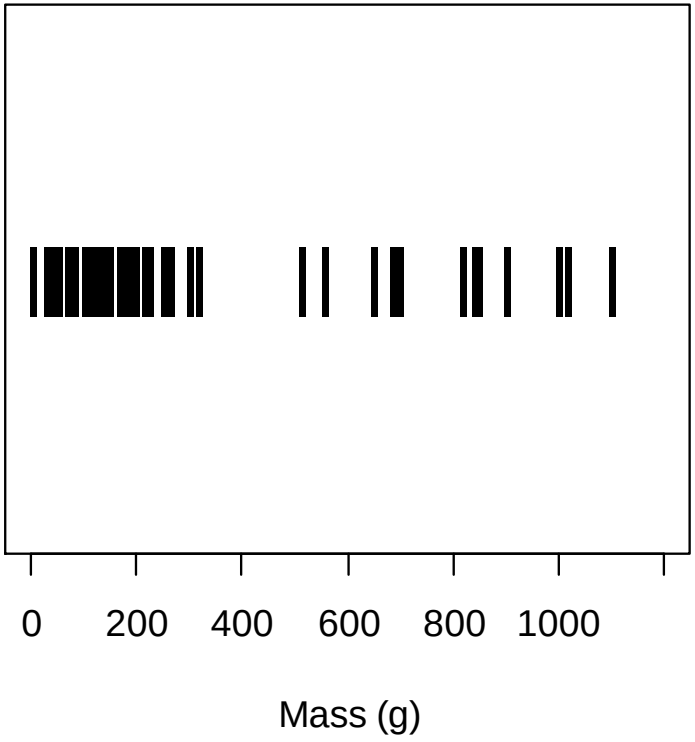
Jittered Dot Plot

```
stripchart(mass,method="jitter",jitt
er=0.2,xlab="Mass
(g)",pch=21,cex=0.5,
xlim=c(0.,1200.))
```



Barcode Plot

```
stripchart(mass,method="overplot",xlab="Mass (g)",pch="|",cex=2,xlim=c(0.,1200.))
```



The Stemplot

- or the stem-and-leaf plot
- depicting distributions of a numerical variable
- Features:
  - For data sets of reasonable size, each observed value of the quantitative variable can be listed thereby giving a complete picture of all the individual data points as well as a picture of the distribution

- Features (continued):
  - The stem can be prepared quickly and easily by hand for small data sets
  - It is useful for hand calculation of quantiles and thus is useful in preparing another well-known EDA plot—the boxplot and even a histogram
- Function : stem()

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

 0 | 134578899
 1 | 01122233345555789
 2 | 0235567
 3 | 002
 4 |
 5 | 16
 6 | 599
 7 | 00
 8 | 2245
 9 | 00
10 | 0002
11 | 00
```

```
1 | 2: represents 120
 leaf unit: 10
         n: 56

   9      0 | 034577888
  28      1 | 01112223333344557889
  (6)     2 | 125566
  22      3 | 002
          4 |
  19      5 | 15
  17      6 | 589
  14      7 | 00
  12      8 | 2245
   8      9 | 00
   6     10 | 0001
   2     11 | 00
```

Praktek di R

Stem Plot

```
plot.new()
sink("test")
stem(mass,scale=2)
sink()
zz<-readLines("test")
text(rep(0,length(zz)),sort(1:length(zz),dec
reasing=TRUE)/(length(zz)),zz,pos=4,font=
1,offset=0)
```

```
The decimal point is 2 digit(s) to the righ

 0 | 134578899
 1 | 0112223333345555789
 2 | 0235567
 3 | 002
 4 |
 5 | 16
 6 | 599
 7 | 00
 8 | 2245
 9 | 00
10 | 0002
11 | 00
```

The Boxplot

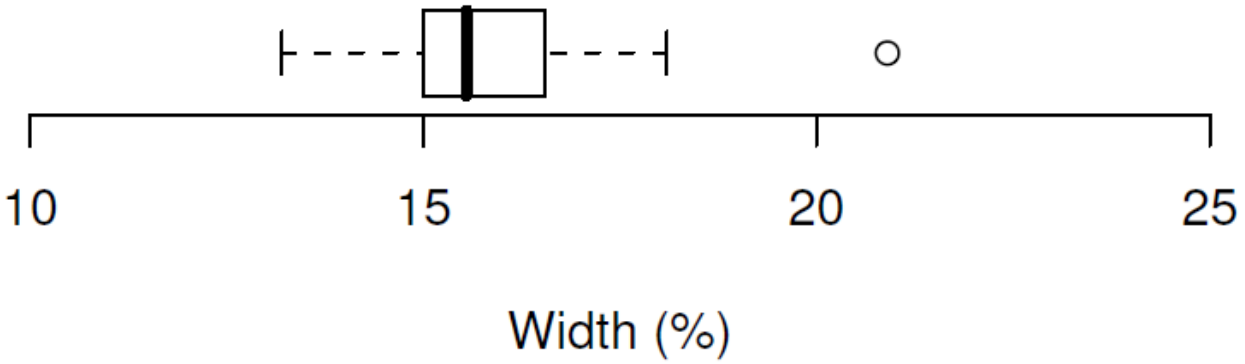
- box-and-whisker plot
- box-whisker plot
- the boxplot require the concept of a quantile
- Function : boxplot()

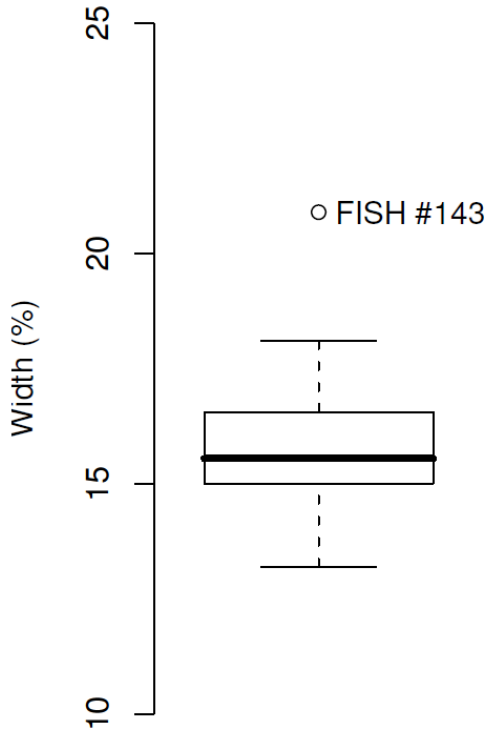
The plotting convention for the quantile boxplot consists of:

- a rectangular box with edges determined by the lower and upper quartiles;
- the median denoted as a line segment splitting the rectangular box into two adjoining boxes;
- a whisker (that is, a line segment) from Q1 to the minimum; and
- a whisker from Q3 to the maximum.

The plotting convention for the outlier boxplot consists of:

- a rectangular box with edges determined by the lower and upper quartiles;
- the median denoted as a line segment splitting the rectangular box into two adjoining boxes;
- whiskers ending at the adjacent values; and
- open circles to represent points lying beyond the inner fences.


10
15
20
25
Width (%)

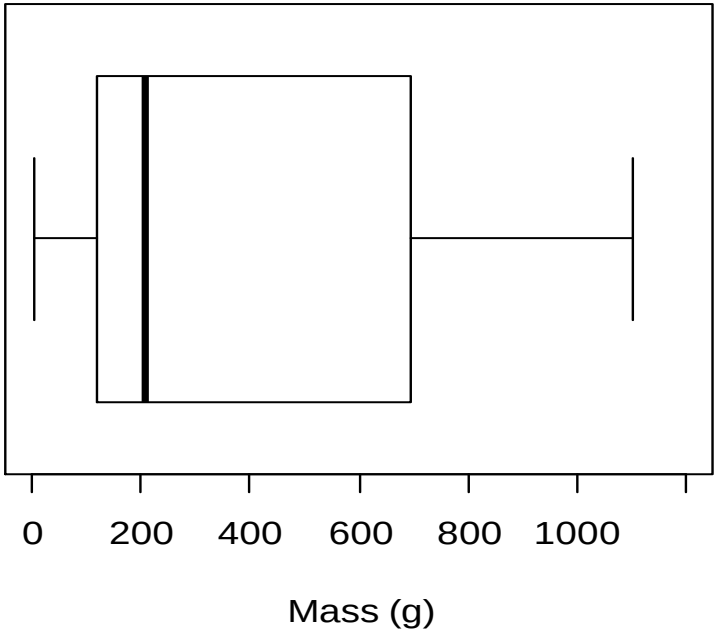

Width (%)
10
15
20
25
FISH #143

Praktek di R

Boxplot-Tipe 1

Masih menggunakan data "mass"

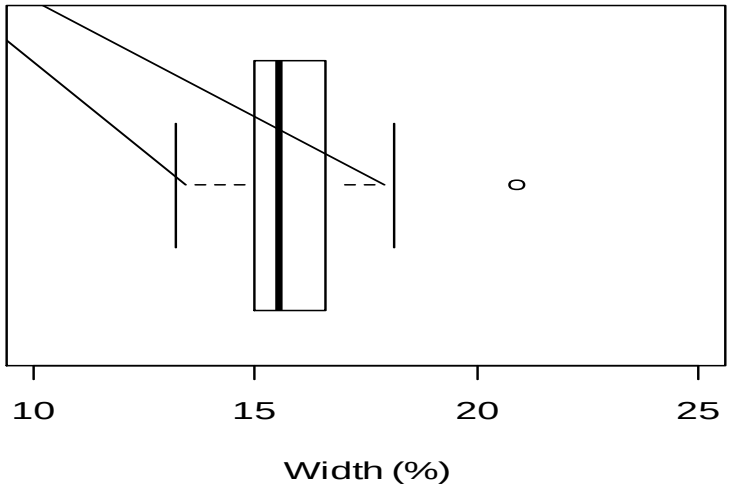
```
boxplot(mass,range=1.5,horizontal
=TRUE,xlab="Mass
(g)",pars=list(boxwex=1.5),
lty=1,ylim=c(0.,1200.),yaxp=c(0.,12
00.,6),outline=FALSE)
```



Boxplot-Tipe 2

Menggunakan data "width"

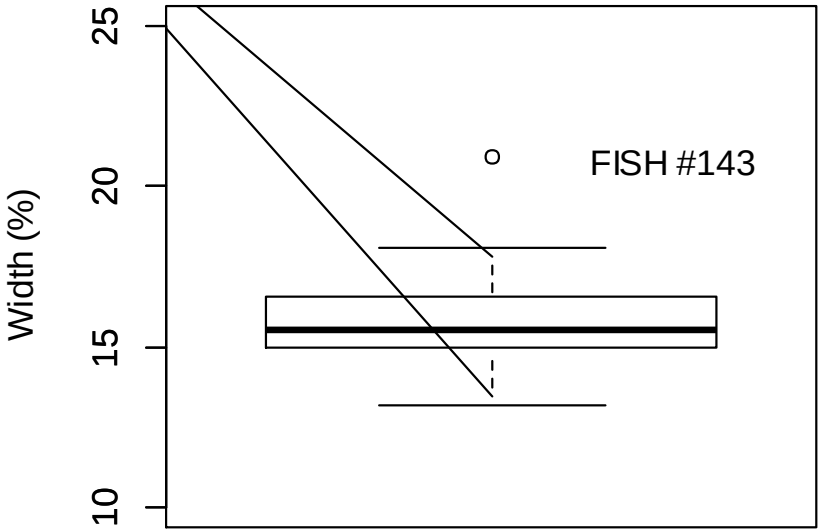
```
width<-
c(16.0,13.6,15.2,15.3,15.9,17.3,16.1,15.1,14.6,
13.2,
15.8,14.7,16.3,15.5,14.5,15.0,15.0,15.0,17.0,1
5.1,15.1,15.0,
14.8,14.9,14.6,15.0,15.9,13.9,15.7,14.8,17.9,1
5.0,15.0,15.8,
14.3,15.4,15.1,17.7,17.5,20.9,17.6,17.6,15.9,1
6.2,18.1,14.5,
17.8,16.8,17.0,17.6,15.6,15.4,16.1,16.3,17.7,1
6.3)
boxplot(width,range=1.5,horizontal=TRUE,pars
=list(boxwex=1.5),xlab="Width (%)",
cex=1.0,ylim=c(10.,25.),yaxp=c(10.,25.,3),outli
ne=TRUE)
```



Boxplot-Tipe 3

Menggunakan data "width"

```
boxplot(width,range=1.5,horizontal=FALSE,p
ars=list(boxwex=1.5),ylab="Width (%)",
cex=1.0,ylim=c(10.,25.),yaxp=c(10.,25.,3),ou
tline=TRUE,
xaxt="n",mfg=c(3,1))
#
text(1.3,20.9,"FISH #143")
#
axis(2,c(10.,15.,20.,25))
```



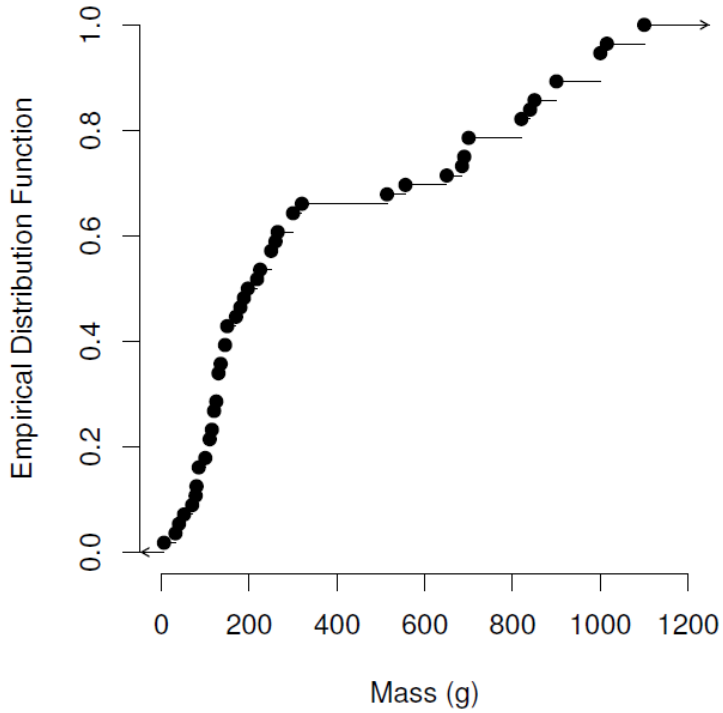
The EDF Plot

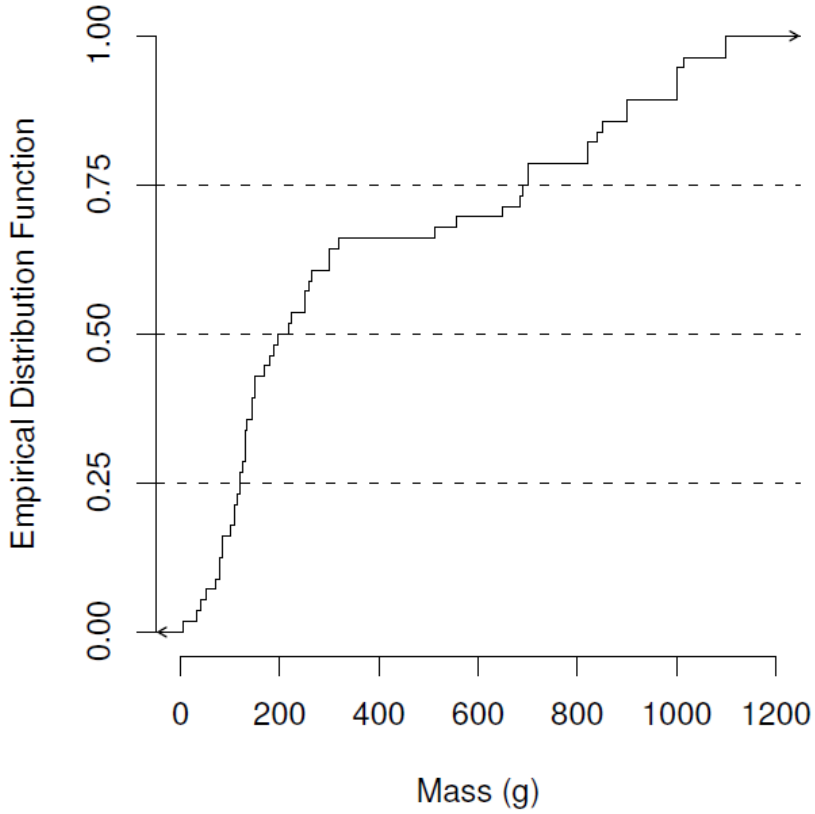
- A discrete approximation to the distribution of the random variable X is given by the empirical distribution function (EDF):

$$S_n(x) = \begin{cases} 0 & \text{if } x < x_{(i)}, \\ \frac{i}{n} & \text{if } x_{(i)} \leq x < x_{(i+1)}, \\ 1 & \text{if } x \geq x_{(n)}. \end{cases}$$

- the empirical distribution function is an approximation to the cumulative distribution function of the random variable X so it is also known as the empirical cumulative distribution function (ecdf)

- plotted along the horizontal axis rather than the vertical axis
- their plot is actually the inverse of the empirical cumulative distribution function’
- Function : ecdf(), plot.stepfun()


Empirical Distribution Function
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
1000
1200
Mass (g)

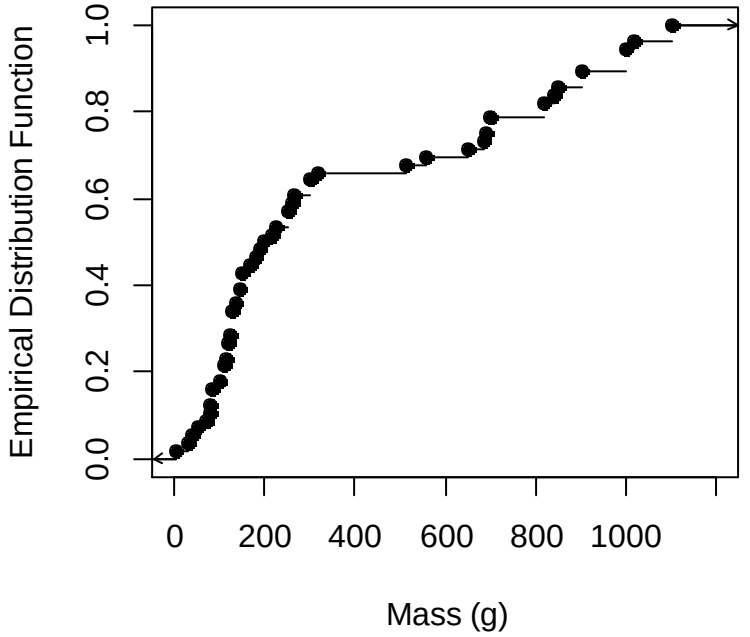

Empirical Distribution Function
0.00
0.25
0.50
0.75
1.00
0
200
400
600
800
1000
1200
Mass (g)

Praktek di R

EDF Plot

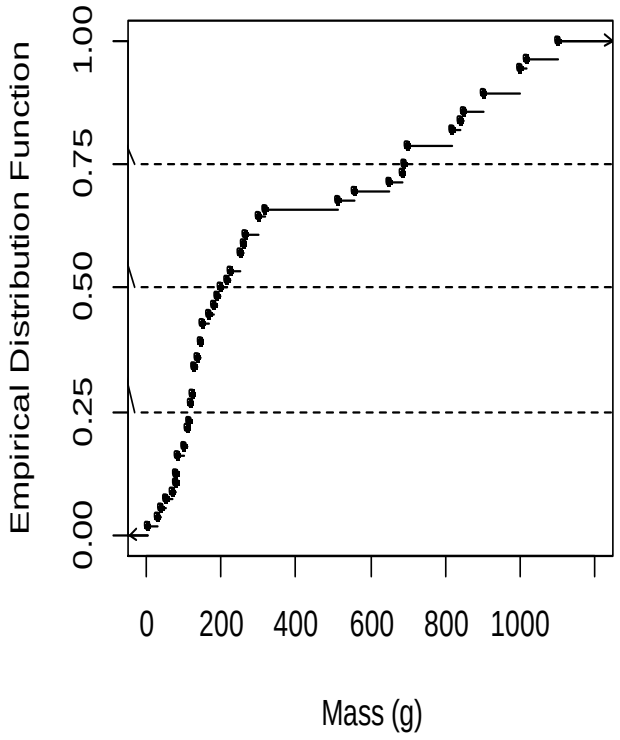
Masih menggunakan data "mass"

```
ecdfmass<-ecdf(mass)
plot.stepfun(ecdfmass,xlab="Mass
(g)",ylab="Empirical Distribution Function",
main=NULL,verticals=FALSE,do.points=TRUE,pc
h=19,xlim=c(0,1200))
#
arrows(max(mass),1.0,1245,1.0,code=2,length
=0.05)
arrows(-
45.,0.0,min(mass),0.0,code=1,length=0.05)
```



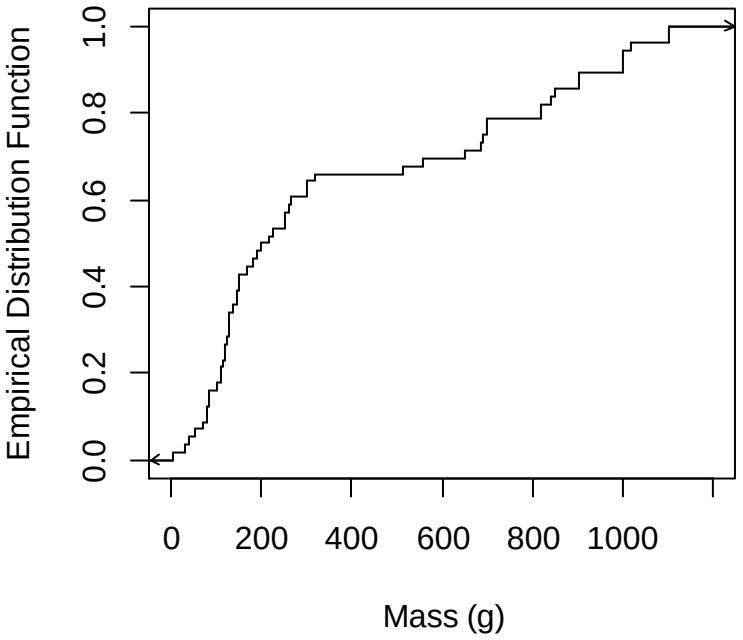
EDF Plot with reference lines for estimation of quartiles

```
ecdfmass<-ecdf(mass)
plot.stepfun(ecdfmass,xlab="Mass
(g)",ylab="Empirical Distribution Function",
main=NULL,verticals=FALSE,do.points=TRUE,pch
=19,cex=0.5,xlim=c(0,1200),
yaxp=c(0.,1.,4),lwd=1.75)
#
arrows(max(mass),1.0,1245,1.0,code=2,length=0
.05)
arrows(-
45.,0.0,min(mass),0.0,code=1,length=0.05)
#
abline(h=c(0.25,0.5,0.75),lty=2)
```



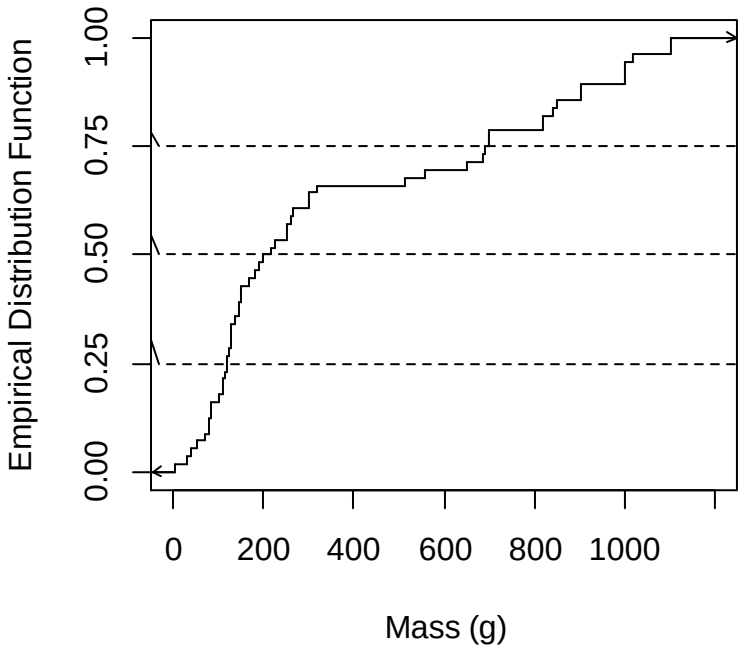
Step plot version of the EDF plot

```
ecdfmass<-ecdf(mass)
plot.stepfun(ecdfmass,xlab="Mass
(g)",ylab="Empirical Distribution
Function",
main=NULL,verticals=TRUE,do.points=FAL
SE,pch=16,xlim=c(0,1200))
#
arrows(max(mass),1.0,1245,1.0,code=2,le
ngth=0.05)
arrows(-
45.,0.0,min(mass),0.0,code=1,length=0.05
)
```



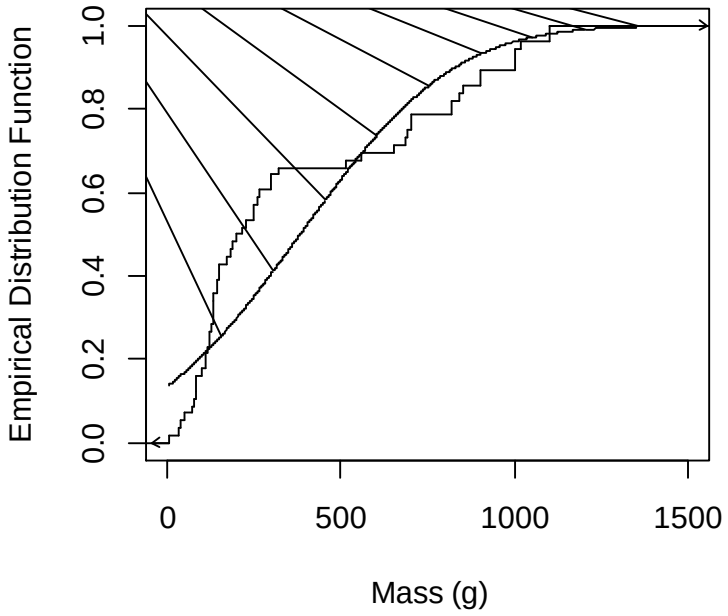
Step plot version of the EDF plot with reference lines for estimation of quartile

```
ecdfmass<-ecdf(mass)
plot.stepfun(ecdfmass,xlab="Mass
(g)",ylab="Empirical Distribution Function",
main=NULL,verticals=TRUE,do.points=FALSE,pc
h=16,xlim=c(0,1200),
yaxp=c(0.,1.,4))
#
arrows(max(mass),1.0,1245,1.0,code=2,length
=0.05)
arrows(-
45.,0.0,min(mass),0.0,code=1,length=0.05)
#
abline(h=c(0.25,0.5,0.75),lty=2)
```



Step plot version of the EDF plot with cumulative normal distribution function added

```
mean_mass<-mean(mass)
sd_mass<-sd(mass)
min_mass<-min(mass)
max_mass<-1500
#
xx<-seq(0,10000,1)*(max_mass-
min_mass)/10000.+min_mass
yy<-pnorm(xx,mean_mass,sd_mass)
#
ecdfmass<-ecdf(mass)
plot.stepfun(ecdfmass,xlab="Mass (g)",ylab="Empirical
Distribution Function",
main=NULL,verticals=TRUE,do.points=FALSE,pch=16,xlim=
c(0,1500))
#
arrows(max(mass),1.0,1555,1.0,code=2,length=0.05)
arrows(-45.,0.0,min(mass),0.0,code=1,length=0.05)
#
lines(xx,yy)
```



Selesai