

Statistical Graphs Using R

Jing Li

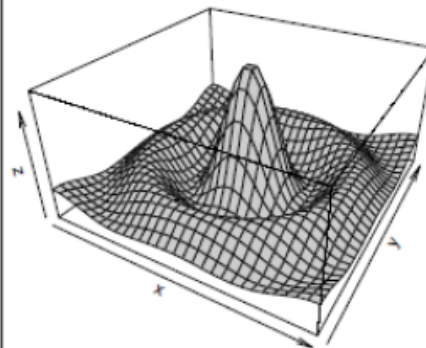
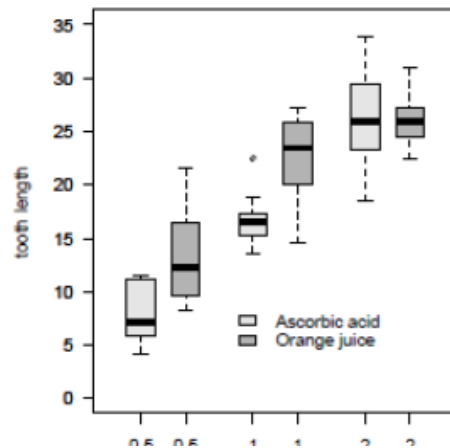
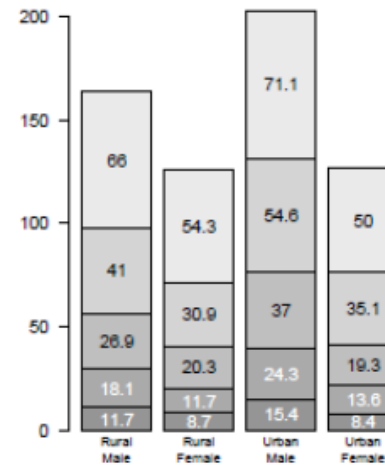
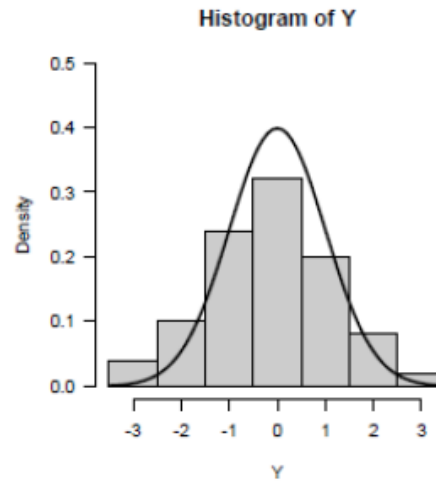
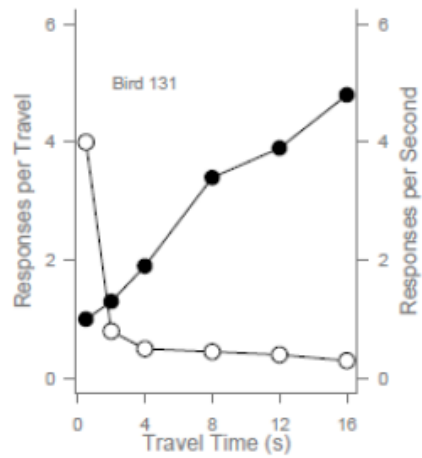
jing.li@sjtu.edu.cn

<http://cbb.sjtu.edu.cn/~jingli/>

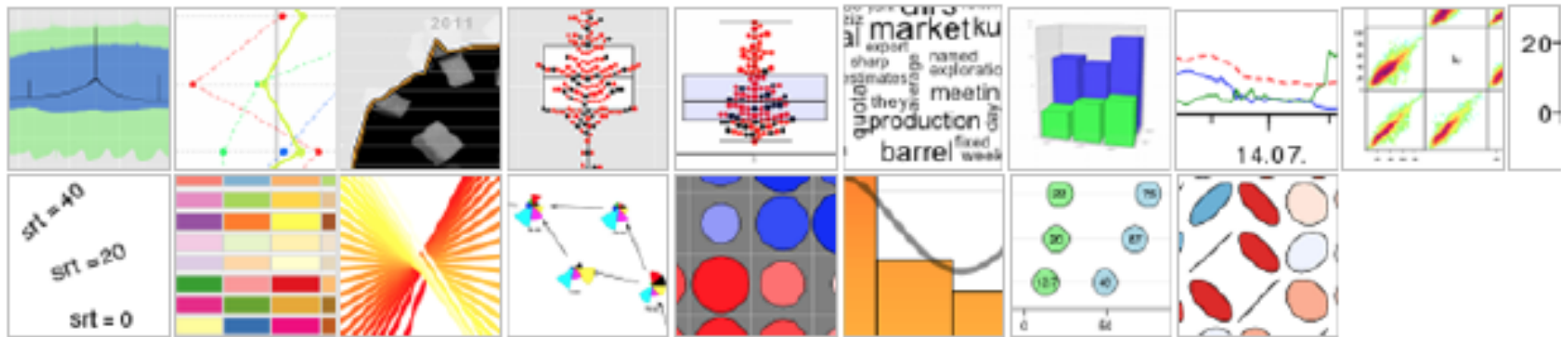
Dept of Bioinformatics & Biostatistics, SJTU



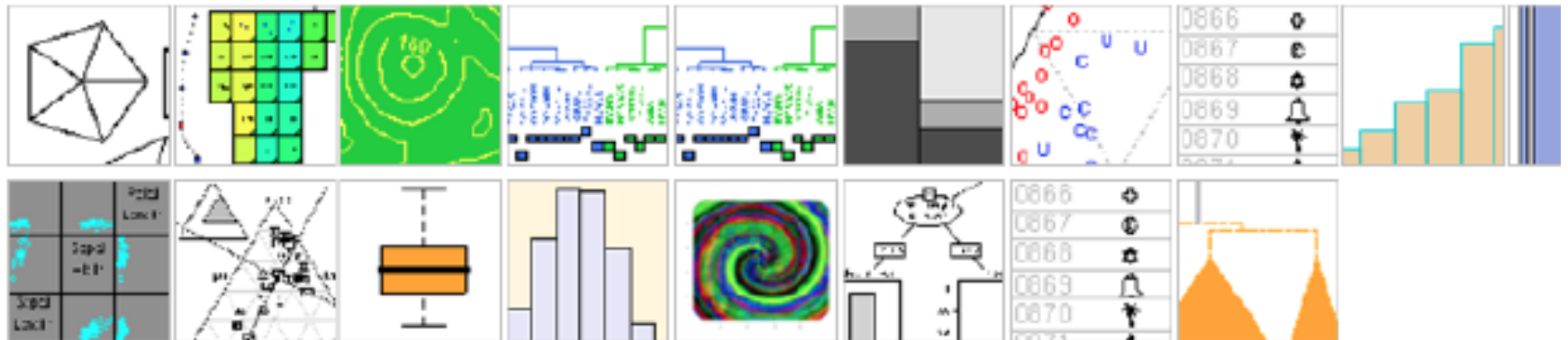
Graphs using R



More examples ...



tries

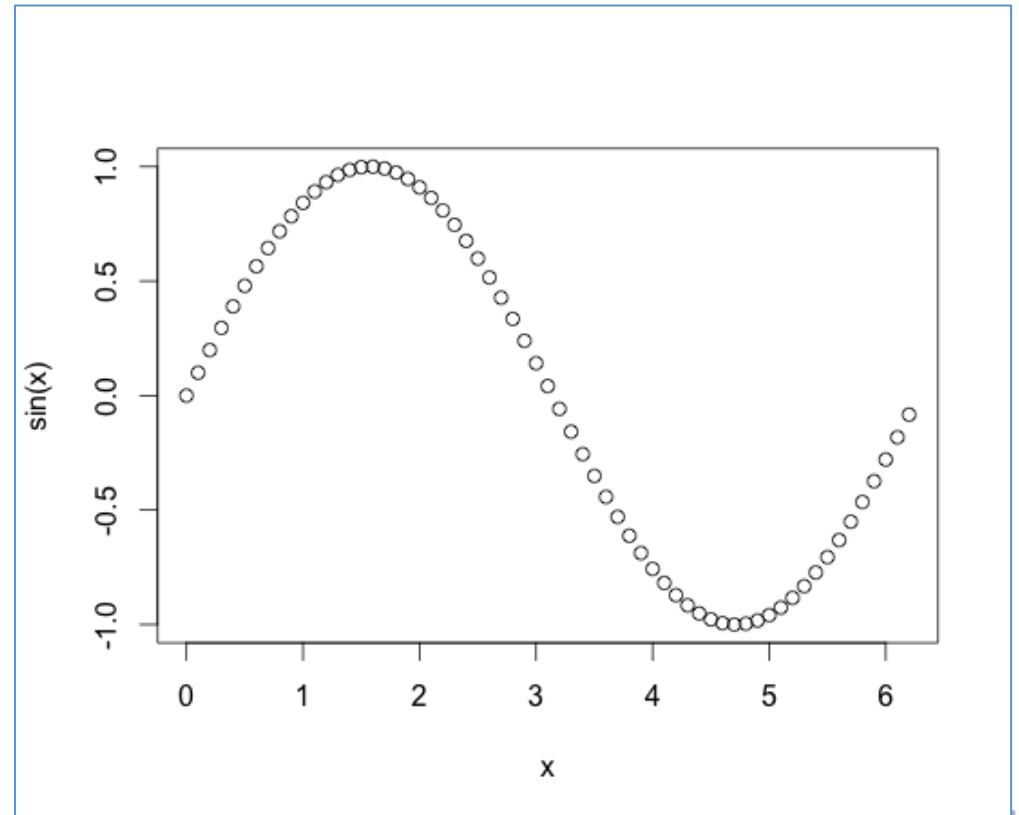


<http://addictedtor.free.fr/graphiques/>



Simplest example

- `x<-seq(0,2*pi,0.1)`
`plot(x,sin(x))`
`plot(x,cos(x))`



Import data

```
>mydata<-read.csv(file="C\\librarian.csv", header=T)
```

```
>head (mydata, 2)
```

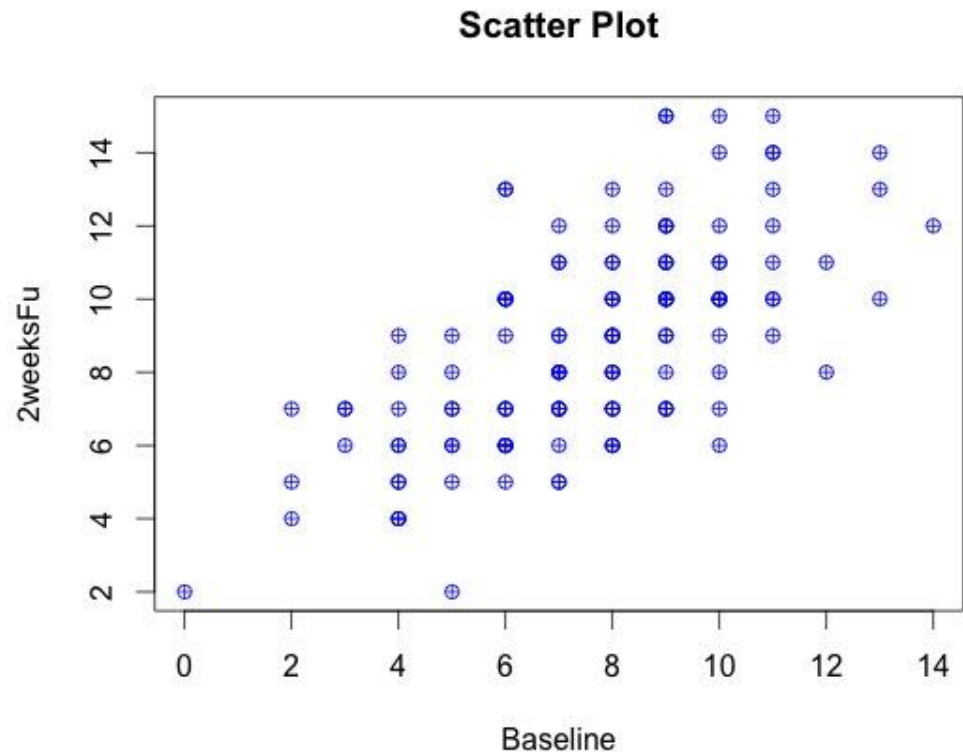
```
>dim(mydata)
```

```
>names(mydata)
```



Scatter plot

```
plot(mydata$base_score,mydata$score1, xlab="Baseline",  
ylab="2weeksFu", main="Scatter Plot", col="blue", pch=10)
```



col specify a color
xlab X axis label
pch specify a symbol
main the main title



plot

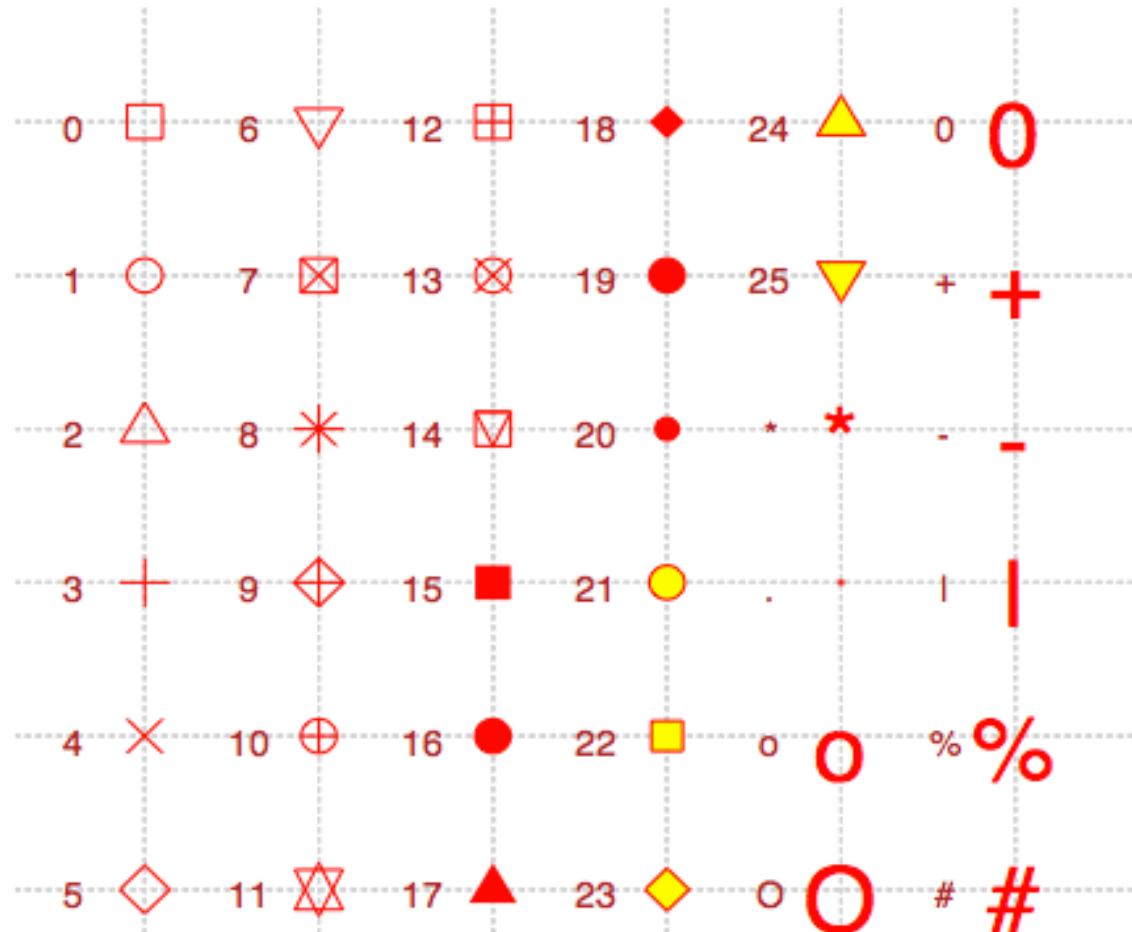
plot(x, y, parameters...)

参数	描述
axes=TRUE	如果是FALSE，不绘制轴与边框
type="p"	指定图形的类型，"p": 点，"l": 线，"b": 点连线，"o": 同上，但是线在点上，"h": 垂直线，"s": 阶梯式，垂直线顶端显示数据，"S": 同上，但是在垂直线底端显示数据
xlim=, ylim=	指定轴的上下限，例如xlim=c(1, 10)或者xlim=range(x)
xlab=, ylab=	坐标轴的标签，必须是字符型值
main=""	主标题，必须是字符型值
sub=""	副标题（用小字体）
pch=12	改变绘图字符
col="yellow"	更改绘图字符颜色，调用colors()查看所有颜色
bg="red"	更改绘图字符内部颜色（仅对pch=21-25的绘图字符有效）



Plot symbols in R

plot symbols: points(..., pch = *, cex = 3)



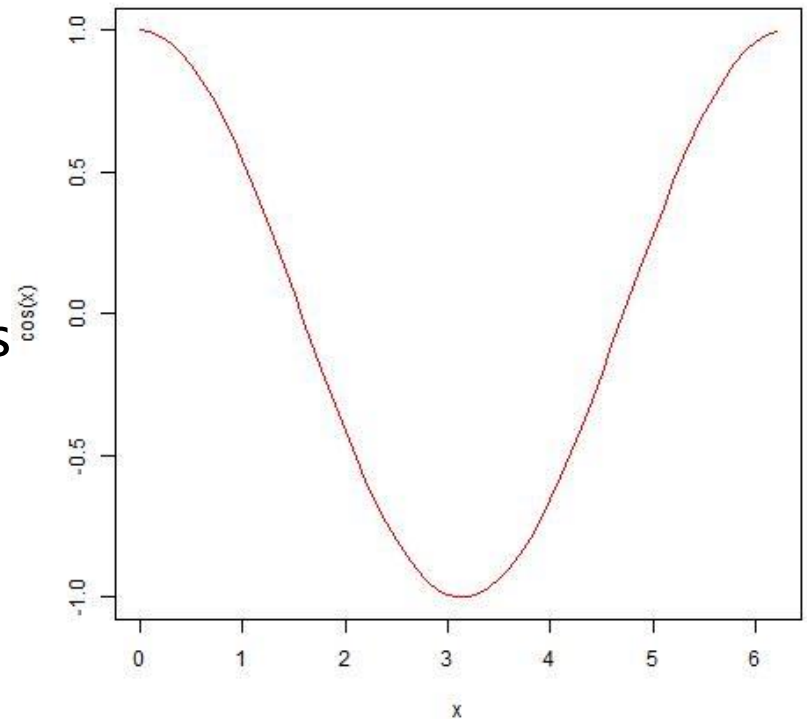
Plot parameters : col & type

col

- "blue"
- "red"
- "yellow"
- ...
- colors()

type

- "p" for points
- "l" for lines
- "o" for both
'overplotted' ,
- "h" for 'histogram'
like (or 'high-
density') vertical lines
- "s" for stair steps,
- "S" for other steps



Example

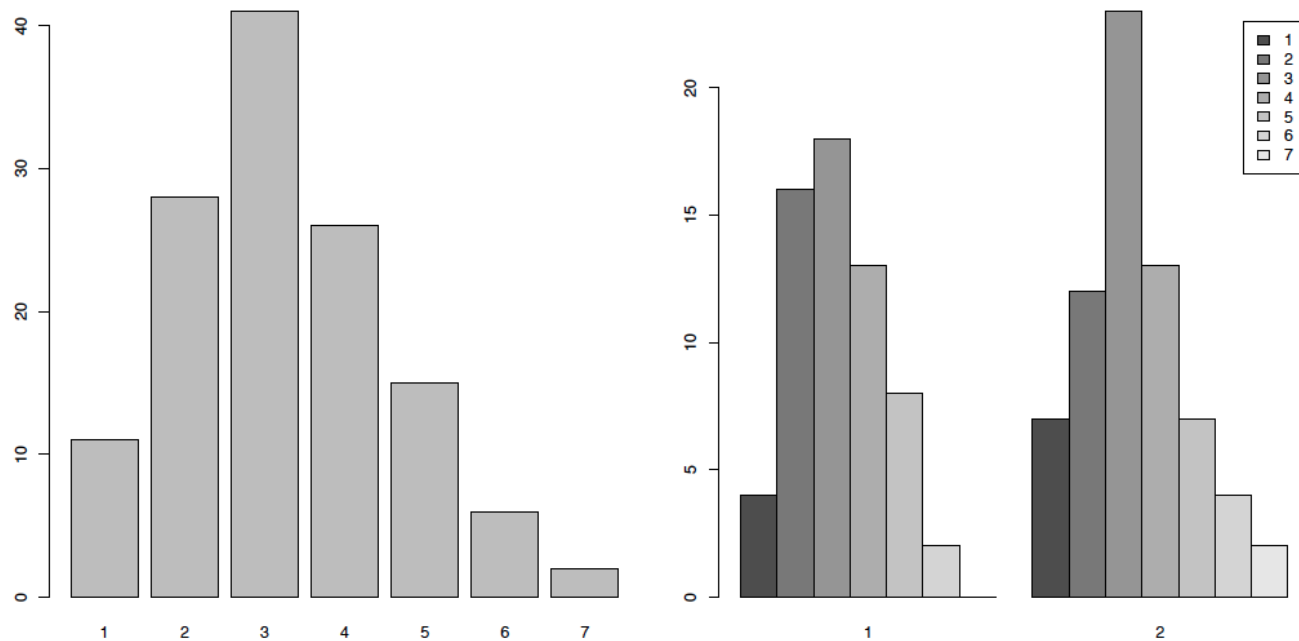
```
x<-seq(0,2*pi,0.1)  
plot(x,cos(x),type="l",col="red")
```



Bar Chart

```
barplot(table(mydata$education))
```

```
barplot(table(mydata$education, mydata$randomization), beside=TRUE,  
legend.text=TRUE )
```



beside FALSE=stacked bars; TRUE=juxtaposed bars

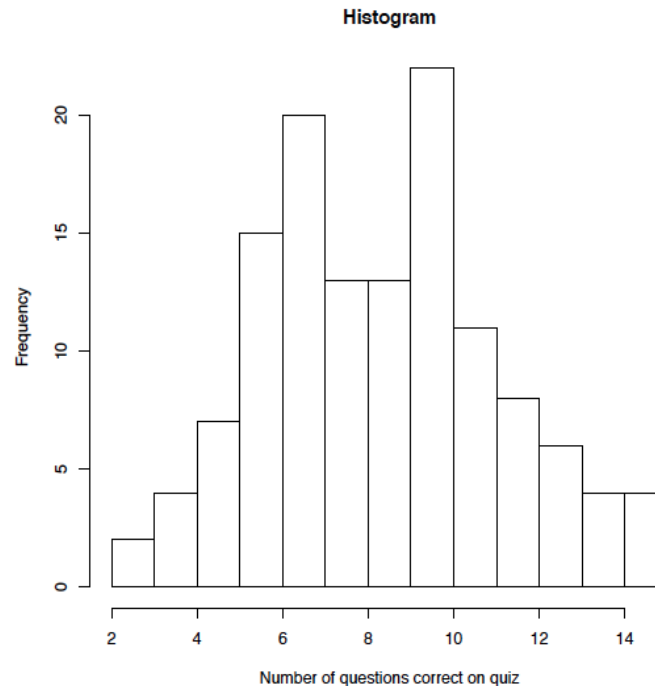
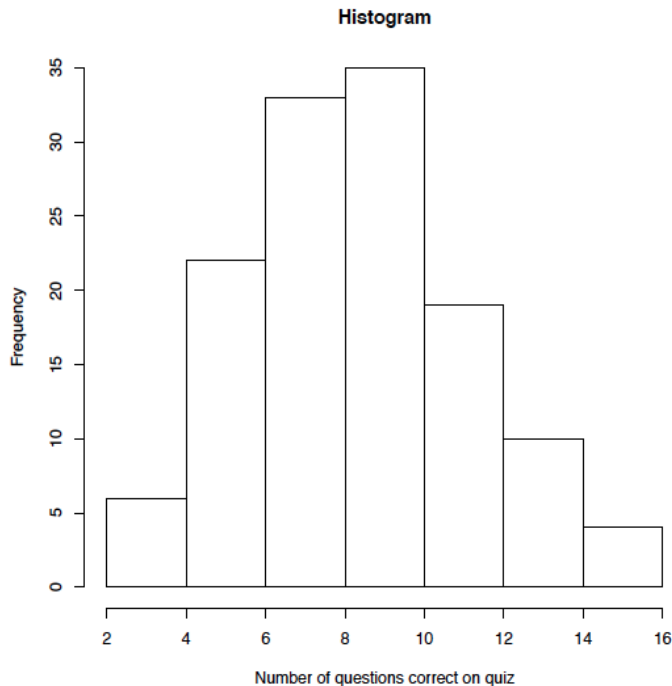
legend.text indicate whether a legend should be included



Histogram

`hist(mydata$score1, main = "Histogram", xlab = "Number of questions correct on quiz")`

`hist(mydata$score1, main = "Histogram", xlab = "Number of questions correct on quiz", breaks = 13)`

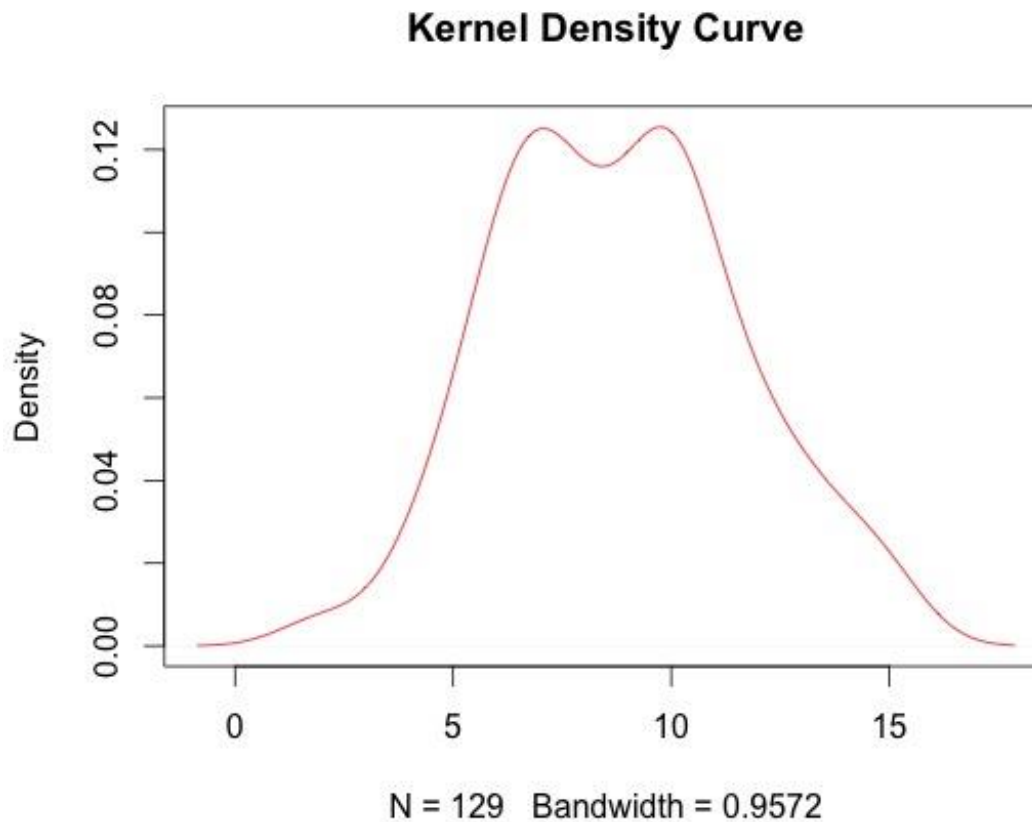


breaks the number of cells for the histogram



Kernel Density Curve

```
plot(density(mydata$score1),main="Kernel Density Curve", col="red")
```



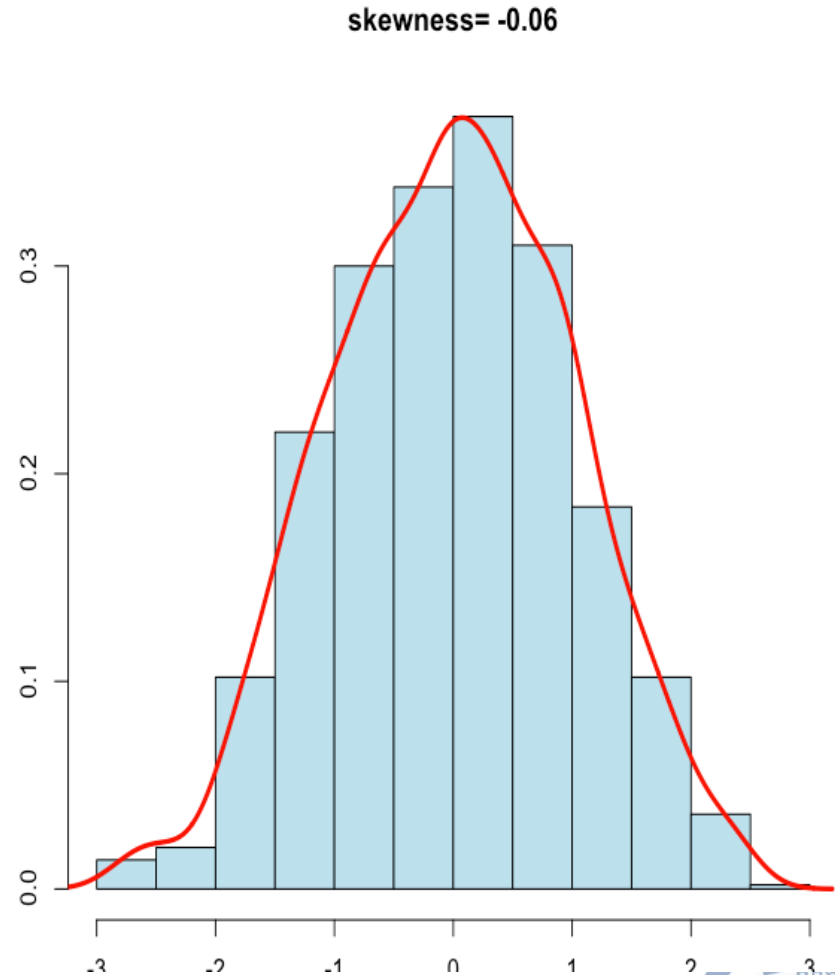
Histogram+ Density curve

- **hist(x)**

```
hist(x, col="light blue",  
probability=TRUE,  
main=paste("skewness=",  
round(skewness(x), digits=2)),  
xlab="", ylab="")
```

- **lines (density (x))**

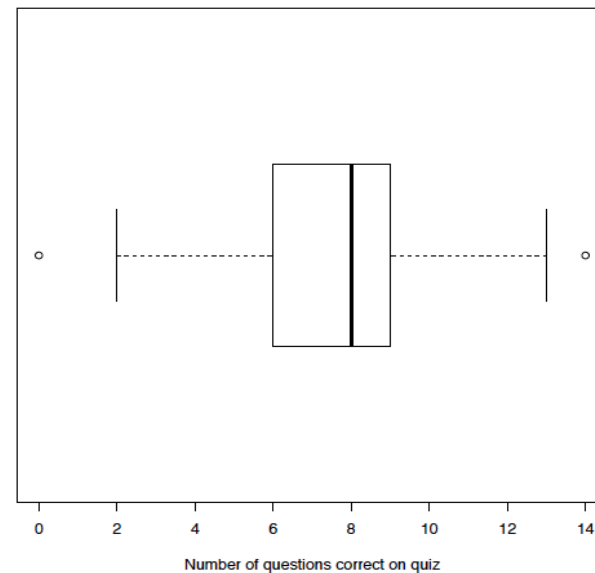
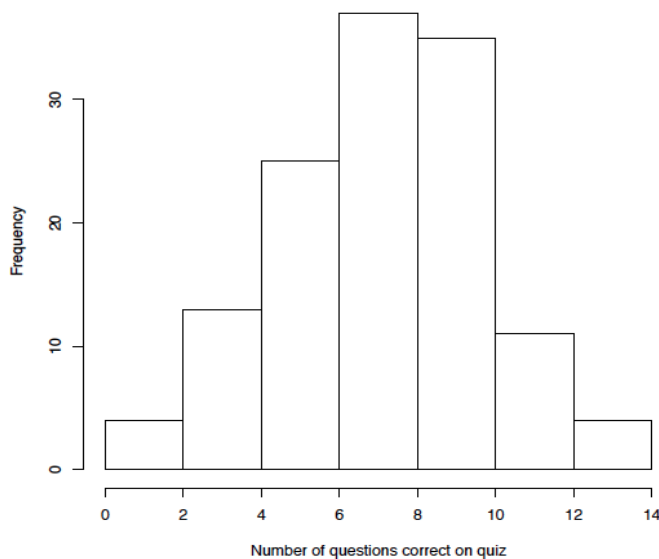
```
lines(density(x), col="red", lwd=3)
```



Boxplot

```
hist(mydata$base_score, main = "", xlab = "Number of questions correct on quiz")
```

```
boxplot(mydata$base_score, horizontal = T, xlab = "Number of questions correct on quiz")
```



horizontal indicate if the boxplots should be horizontal

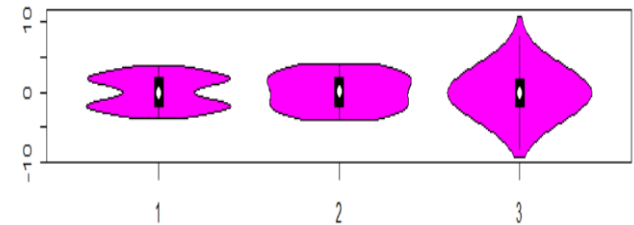


Violin plot

```
install.packages("vioplot")
```

```
library(vioplot)
```

```
vioplot(x1,x2,....,names=,col=)
```



Output of graph

To redirect graphic output use one of the following functions.
Use `dev.off()` to return output to the terminal.

Function	Output to
<code>pdf("mygraph.pdf")</code>	pdf file
<code>win.metafile("mygraph.wmf")</code>	windows metafile
<code>png("mygraph.png")</code>	png file
<code>jpeg("mygraph.jpg")</code>	jpeg file
<code>bmp("mygraph.bmp")</code>	bmp file
<code>postscript("mygraph.ps")</code>	postscript file

Redirecting Graphs

```
# example - output graph to jpeg file  
jpeg("c:/mygraphs/myplot.jpg")  
plot(x)  
dev.off()
```



More functions

命令	描述
<code>points(x, y)</code>	添加点（可以使用选项 <code>type=</code> ）
<code>lines(x, y)</code>	同上，但是添加线
<code>text(x, y, labels,...)</code>	在(x,y)处添加用labels指定的文字；典型的用法是： <code>plot(x, y,type="n"); text(x, y, names)</code>
<code>segments(x0, y0,x1, y1)</code>	从(x0,y0)各点到(x1,y1)各点画线段
<code>arrows(x0, y0,x1, y1, angle= 30,code=2)</code>	同上但加画箭头，如果 <code>code=2</code> 则在各(x0,y0)处画箭头，如果 <code>code=1</code> 则在各(x1,y1)处画箭头，如果 <code>code=3</code> 则在两端都画箭头； <code>angle</code> 控制箭头轴到箭头边的角度
<code>abline(a,b)</code>	绘制斜率为b和截距为a的直线
<code>rect(x1, y1, x2,y2)</code>	绘制长方形，(x1, y1)为左下角，(x2,y2)为右上角
<code>polygon(x, y)</code>	绘制连接各x,y坐标确定的点的多边形
<code>legend(x, y,legend)</code>	在点(x,y)处添加图例，说明内容由legend给定
<code>title()</code>	添加标题，也可添加一个副标题
<code>axis(side, vect)</code>	画坐标轴， <code>side=1</code> 时画在下边， <code>side=2</code> 时画在左边， <code>side=3</code> 时画在上边， <code>side=4</code> 时画在右边。
<code>box()</code>	在当前的图上加上边框



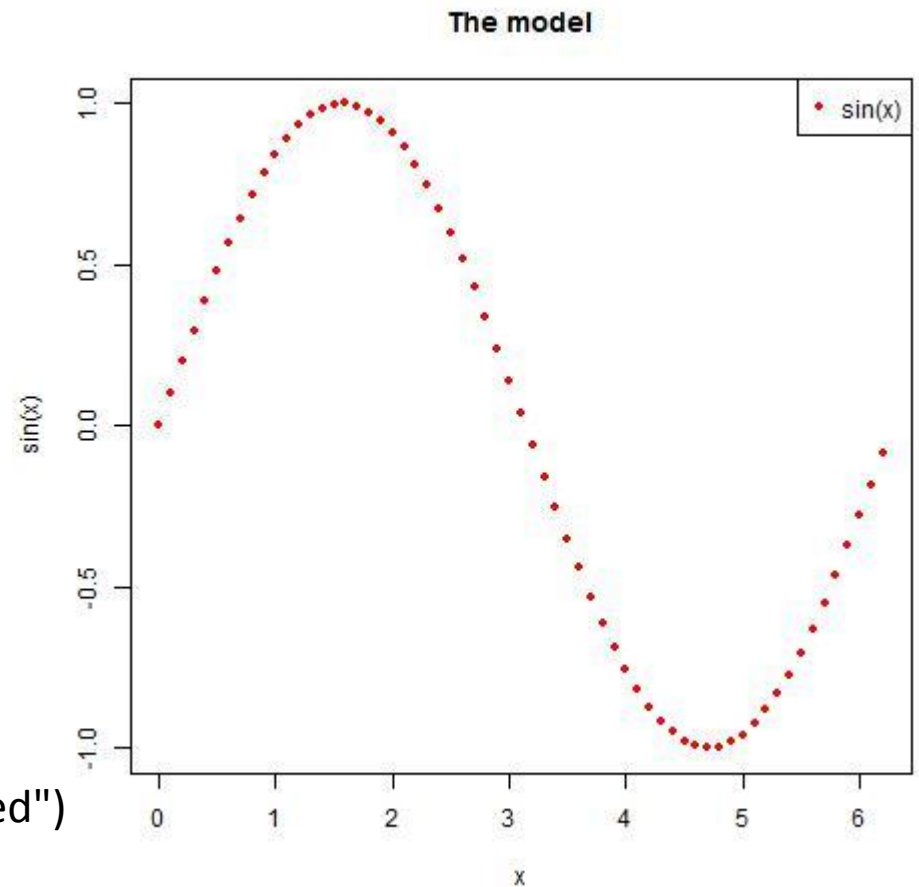
Add legend & title

```
legend(x, y = NULL, legend, fill = NULL,  
col = par("col"), border="black", lty, lwd,  
pch...)
```

```
title(main = NULL, sub = NULL, xlab =  
NULL, ylab = NULL, line = NA, outer =  
FALSE, ...)
```

Example

```
plot(x,sin(x), col="red", pch=20)  
legend("topright", "sin(x)", pch = 20, col="red")  
title("The model")
```

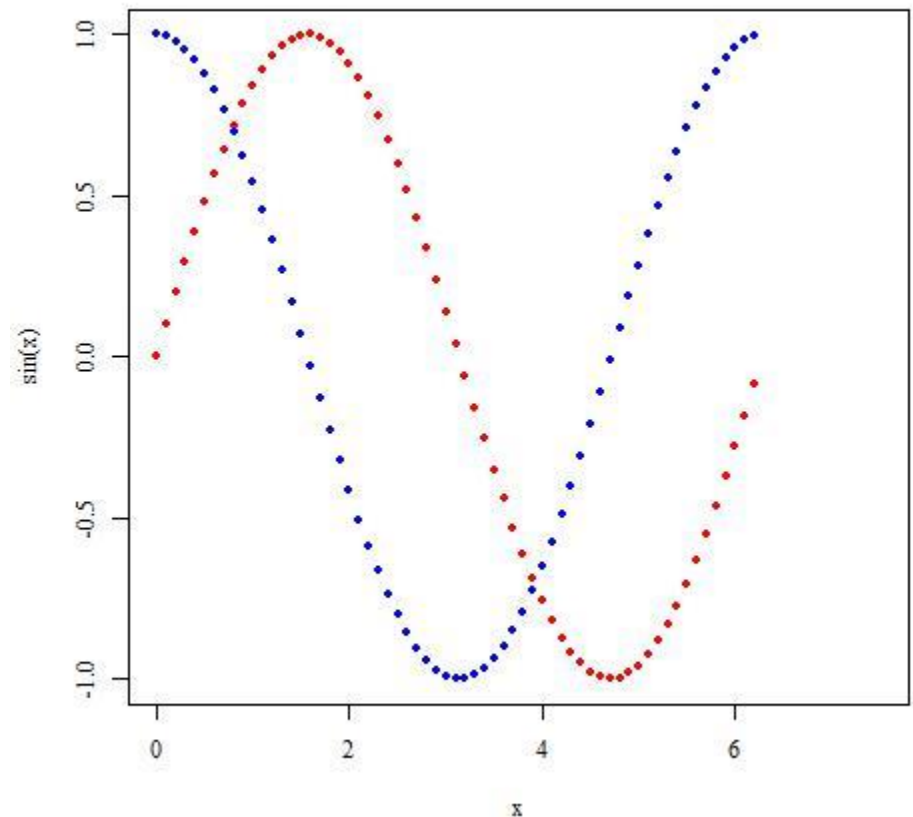


Add point / line

- `lines(x, y, ...)` : 在图像增加线
- `points(x, y, ...)` : 在图像上增加点
- `rect(x1, y1, x2,y2)` : 绘制长方形, $(x1, y1)$ 为左下角, $(x2,y2)$ 为右上角
- `polygon(x, y)` : 绘制连接各 x,y 坐标确定的点的多边形

实例演示 :

```
x<-seq(0,2*pi,0.1)
plot(x,sin(x), col="red", pch=20, family =
"serif", xlim = c(0, 7.5))
for(x in seq(0,2*pi,0.1)){
points(x,cos(x),col="blue",pch=20, family =
"serif")
}
```



Options

参数	描述
adj	控制关于文字的对齐方式，0是左对齐，0.5是居中对齐，1是右对齐，值>1时对齐位置在文本右边的地方，取负值时对齐位置在文本左边的地方
bg	指定背景色（例如bg="red", bg="blue"; 用colors()可以显示657种可用的颜色名）
bty	控制图形边框形状，可用的值为:"o", "l", "7", "c", "u" 和 "]"（边框和字符的外表相像）；如果bty="n"则不绘制边框
cex	控制缺省状态下符号和文字大小的值；另外，cex.axis控制坐标轴刻度数字大小，cex.lab控制坐标轴标签文字大小，cex.main控制标题文字大小，cex.sub控制副标题文字大小
font	控制文字字体的整数（1: 正常，2: 斜体，3: 粗体，4: 粗斜体）；和cex类似，还可用: font.axis, font.lab, font.main, font.sub
las	控制坐标轴刻度数字标记方向的整数（0: 平行于轴，1: 横排，2: 垂直于轴，3: 竖排）
lty	控制连线的线型，可以是整数（1: 实线，2: 虚线，3: 点线，4: 点虚线，5: 长虚线，6: 双虚线）
lwd	控制连线宽度的数字
mar	控制图形边空的有4个值的向量c(bottom, left, top, right), 缺省值为c(5.1, 4.1, 4.1, 2.1)mfcollc(nr,nc)的向量，分割绘图窗口为nr行nc列的矩阵布局，按列次序使用各子窗口
mfcoll	mfcollc(nr,nc)的向量，分割绘图窗口为nr行nc列的矩阵布局，按列次序使用各子窗口
mfrow	同上，但是按行次序使用各子窗口
pch	控制符号的类型，可以是1到25的整数，也可以是""里的单个字符
ps	控制文字大小的整数，单位为磅(points)

Splitting Graph

```
mat <- matrix(1:4, 2, 2)
```

```
> mat
```

```
[,1] [,2]
```

```
[1,] 1 3
```

```
[2,] 2 4
```

```
> layout(mat)
```

```
> layout.show(4)
```

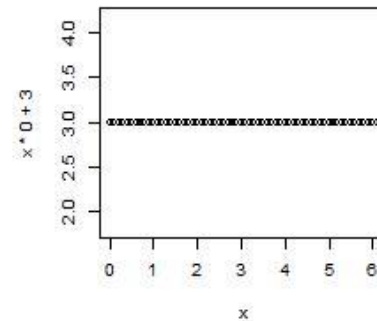
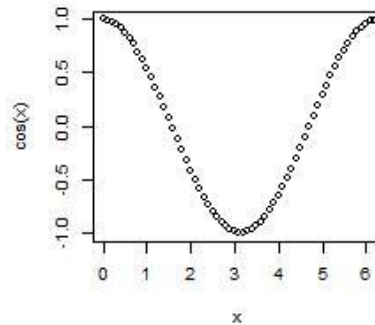
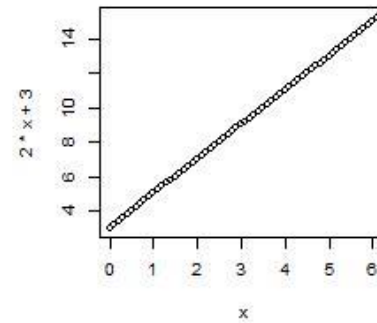
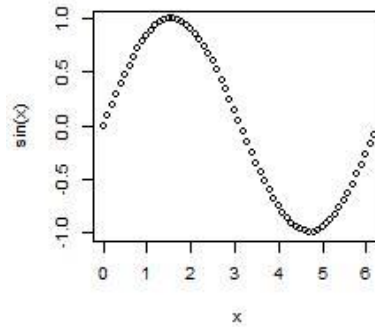
```
x <- seq(0, 2*pi, 0.1)
```

```
plot(x, sin(x))
```

```
plot(x, cos(x))
```

```
plot(x, 2*x+3)
```

```
plot(x, x*0+3)
```



Exercise (1)

```
>x <- stats::runif(12); y <- stats::rnorm(12)
> i <- order(x,y); x <- x[i]; y <- y[i]
> plot(x, y, main="arrows(.) and segments(.)")
> ## draw arrows from point to point :
> s <- seq(length(x)-1)# one shorter than data
> arrows(x[s], y[s], x[s+1], y[s+1], col= 1:3)
> s <- s[-length(s)]
> segments(x[s], y[s], x[s+2], y[s+2], col= 'pink')
```



Exercise (2)

- Draw a boxplot for `chew_score` in `mydata` (`librarian.csv`)
- Draw a boxplot plot for `age` stratified by `randomization`
- Draw a histogram with density line for `score2`
- Draw a scatter plot for `score1` versus `score2`

* Each graph has different **color, type** (point, line)

** Shown your own style, save your graphs as jpeg file



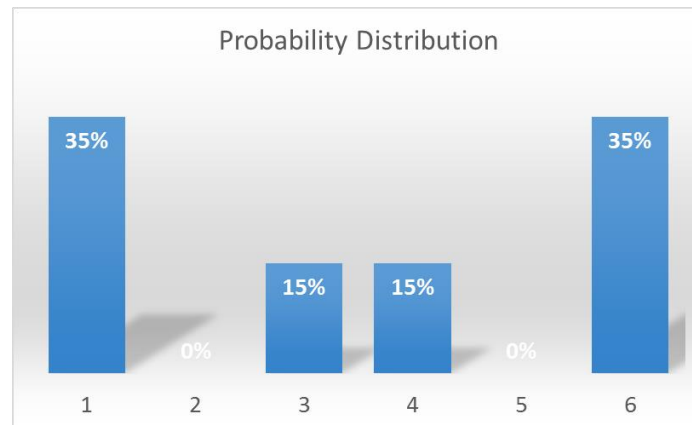
Exercise (3)

- Based on the data in Lab 1-exercise2, display the sample means of each scenario with boxplot in one figure file.



Exercise (4)

- We have a discrete probability distribution as following graph:



- (1) With sample size $n=4$, please take samples from this distribution randomly, and repeat 10,000 times. Then display the distribution of sample means and the mean and SD of sample means.
- (2) Perform similar work with sample size $n=20$, and 100.

