海藻数据分析

班级：硕士三班
学号：2120151064
姓名：张露露

数据可视化和摘要

R 是一套完整的数据处理、计算和制图软件系统。其功能包括：数据存储和处理系统；数组运算工具（其向量、矩阵运算方面功能尤其强大）；完整连贯的统计分析工具；优秀的统计制图功能；简便而强大的编程语言：可操纵数据的输入和输出，可实现分支、循环，用户可自定义功能。
本实验使用 R 软件，版本为 3.3.0，系统为 Windows10.

数据摘要

在 R 软件中打开数据，然后输入命令
summary(algae),得以下结果

```
> algae<-read.table('C:/HZTAO/course/dataMining/R/Analysis.txt',header=T,
> summary(algae)
    season       size       speed         mxPH            mnO2
 autumn:40   large :45   high  :84   Min.   :5.600   Min.   : 1.500
 spring:53   medium:84   low   :33   1st Qu.:7.700   1st Qu.: 7.725
 summer:45   small :71   medium:83   Median :8.060   Median : 9.800
 winter:62                           Mean   :8.012   Mean   : 9.118
                                     3rd Qu.:8.400   3rd Qu.:10.800
                                     Max.   :9.700   Max.   :13.400
                                     NA's   :1       NA's   :2
       Cl                NO3              NH4                oPO4
 Min.   :  0.222   Min.   : 0.050   Min.   :    5.00   Min.   :  1.00
 1st Qu.: 10.981   1st Qu.: 1.296   1st Qu.:   38.33   1st Qu.: 15.70
 Median : 32.730   Median : 2.675   Median :  103.17   Median : 40.15
 Mean   : 43.636   Mean   : 3.282   Mean   :  501.30   Mean   : 73.59
 3rd Qu.: 57.824   3rd Qu.: 4.446   3rd Qu.:  226.95   3rd Qu.: 99.33
 Max.   :391.500   Max.   :45.650   Max.   :24064.00   Max.   :564.60
 NA's   :10        NA's   :2        NA's   :2          NA's   :2
```

```
      PO4              Chla              a1               a2
Min.   :  1.00   Min.   :  0.200   Min.   : 0.00   Min.   : 0.000
1st Qu.: 41.38   1st Qu.:  2.000   1st Qu.: 1.50   1st Qu.: 0.000
Median :103.29   Median :  5.475   Median : 6.95   Median : 3.000
Mean   :137.88   Mean   : 13.971   Mean   :16.92   Mean   : 7.458
3rd Qu.:213.75   3rd Qu.: 18.308   3rd Qu.:24.80   3rd Qu.:11.375
Max.   :771.60   Max.   :110.456   Max.   :89.80   Max.   :72.600
NA's   :2        NA's   :12
      a3               a4               a5               a6
Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000
Median : 1.550   Median : 0.000   Median : 1.900   Median : 0.000
Mean   : 4.309   Mean   : 1.992   Mean   : 5.064   Mean   : 5.964
3rd Qu.: 4.925   3rd Qu.: 2.400   3rd Qu.: 7.500   3rd Qu.: 6.925
Max.   :42.800   Max.   :44.600   Max.   :44.400   Max.   :77.600

      a7
Min.   : 0.000
1st Qu.: 0.000
Median : 1.000
Mean   : 2.495
3rd Qu.: 2.400
Max.   :31.600
```

从图中结果可以看出

season 的可能取值的频数

autumn：45　　spring：53　　summer：45　　winter：62

size 的可能取值的频数

large:45　　medium: 84　　small：71

speed 的可能取值的频数：

high：84　　low：33　　medium：83

	min	Q1	median	mean	Q3	max	NA
mxPH	5.600	7.700	8.060	8.012	8.400	9.700	1
mnO2	1.500	7.725	9.800	9.118	10.800	13.400	2
Cl	0.222	10.981	32.730	43.636	57.824	391.500	10
NO3	0.050	1.296	2.675	3.282	4.446	45.650	2
NH4	5.00	38.33	103.17	501.30	226.95	24064.00	2
oPO4	1.00	15.70	40.15.	73.59	99.33	564.60	2
PO4	1.00	41.38	103.29	137.88	213.75	771.60	2
Chla	0.200	2.000	5.475	13.971	18.308	110.456	12
A1	0.00	1.50	6.95	16.92	24.80	89.80	0
A2	0.000	0.000	3.000	7.458	11.375	72.600	0
A3	0.000	0.000	1.550	4.309	4.925	42.800	0
A4	0.000	0.000	0.000	1.992	2.400	44.600	0
A5	0.000	0.000	1.900	5.064	7.500	44.400	0
A6	0.000	0.000	0.000	5.964	6.925	77.600	0
A7	0.000	0.000	1.000	2.495	2.400	31.600	0

数据可视化

mxPH 的直方图并绘制 QQ 图检验是否为正态分布：

绘制直方图

```
> hist(algae$mxPH)
> hist(algae$mnO2)
> hist(algae$Cl)
> hist(algae$NO3)
> hist(algae$NH4)
> hist(algae$oPO4)
> hist(algae$PO4)
> hist(algae$Chla)
```

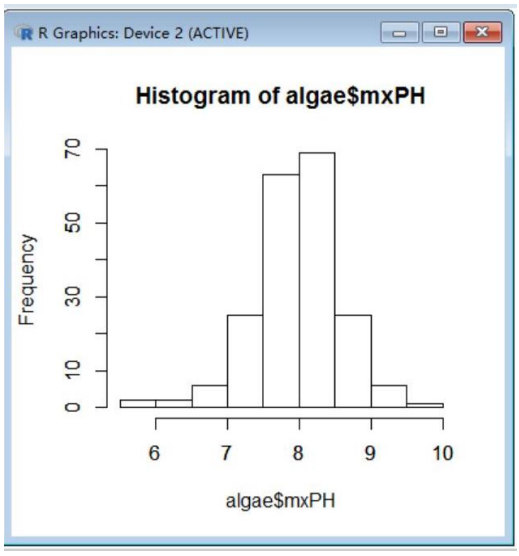
绘制 QQ 图

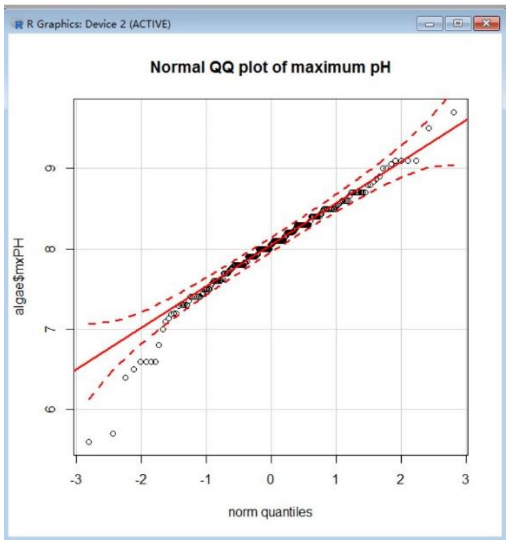
```
> library(car)
> qqPlot(algae$mxPH,main='Normal QQ plot of maximum pH')
> qqPlot(algae$mnO2,main='Normal QQ plot of mnO2')
> qqPlot(algae$ClH,main='Normal QQ plot of Cl')
> qqPlot(algae$NO3,main='Normal QQ plot of NO3')
> qqPlot(algae$NH4,main='Normal QQ plot of NH4')
> qqPlot(algae$oPO4,main='Normal QQ plot of oPO4')
> qqPlot(algae$PO4,main='Normal QQ plot of PO4')
> qqPlot(algae$Chla,main='Normal QQ plot of Chla')
```

得到结果如下图，QQ 图绘制了变量值与正态分布的理论分位数的散点图，同时它给出了正态分布的 95%的置信区间的带状图.

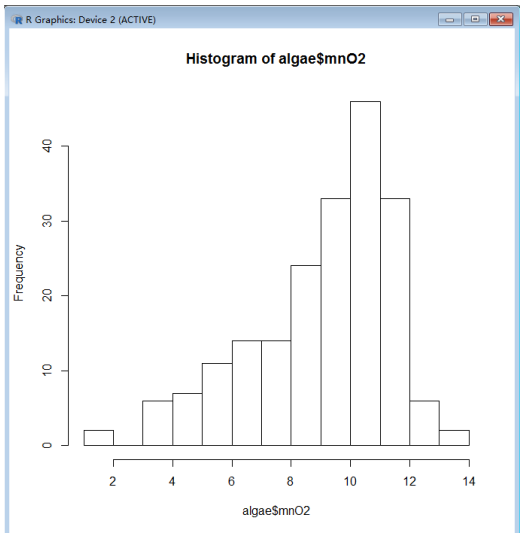
由图可知，变量有几个小的值明显在 95%的置信度区间之外，所以它们都不服从正态分布。

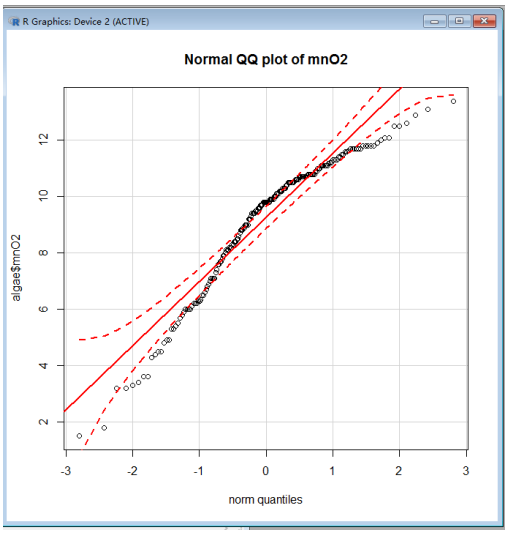
mxPH



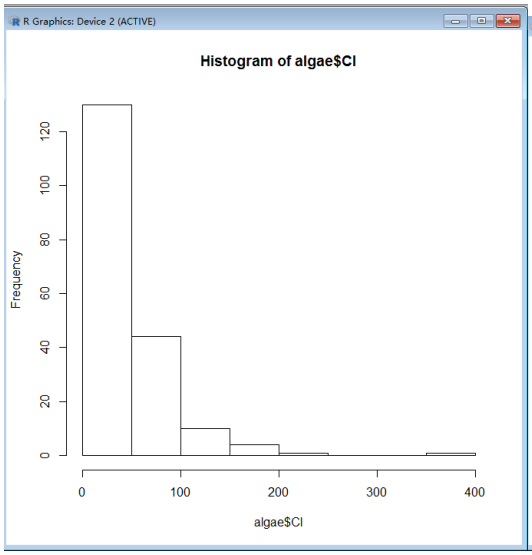


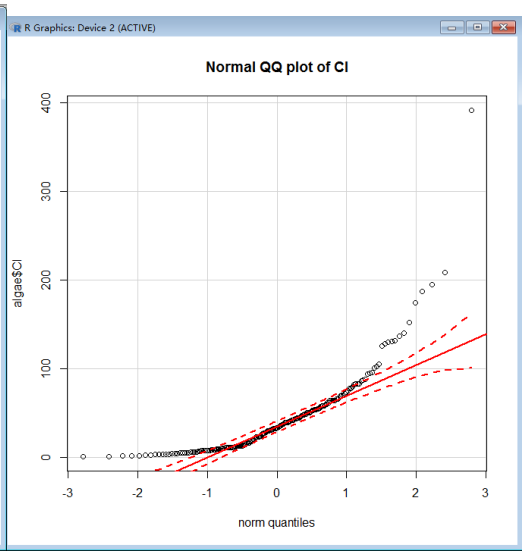
mnO2


R Graphics: Device 2 (ACTIVE)
Histogram of algae$mnO2
Frequency
0
10
20
30
40
2
4
6
8
10
12
14
algae$mnO2

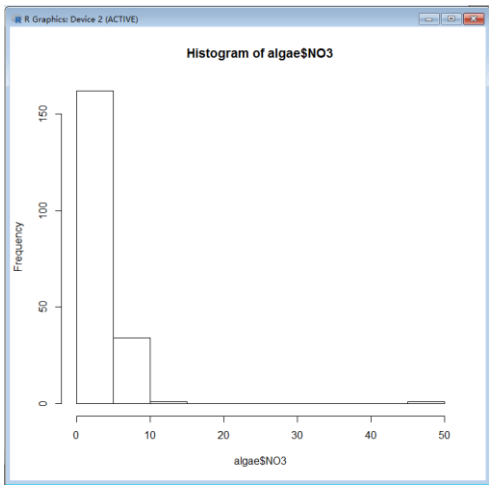

R Graphics: Device 2 (ACTIVE)
Normal QQ plot of mnO2
algae$mnO2
2
4
6
8
10
12
-3
-2
-1
0
1
2
3
norm quantiles

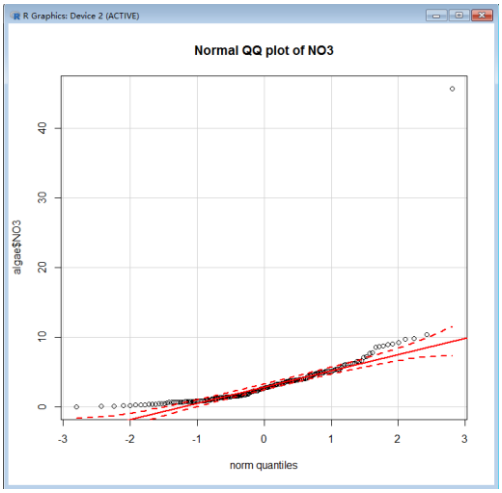
Cl:


R Graphics: Device 2 (ACTIVE)
Histogram of algae$Cl
Frequency
0
20
40
60
80
100
120
0
100
200
300
400
algae$Cl

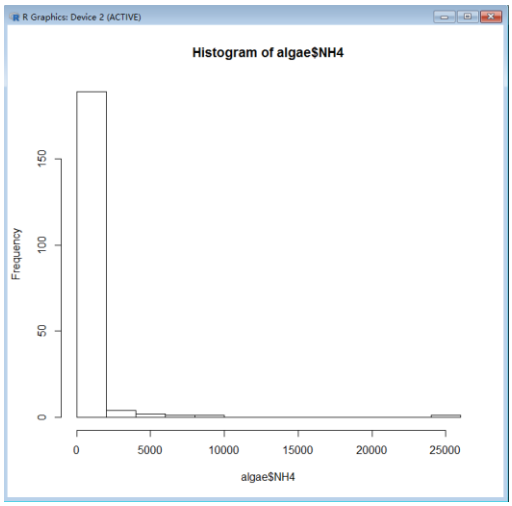

R Graphics: Device 2 (ACTIVE)
Normal QQ plot of Cl
algae$Cl
0
100
200
300
400
-3
-2
-1
0
1
2
3
norm quantiles

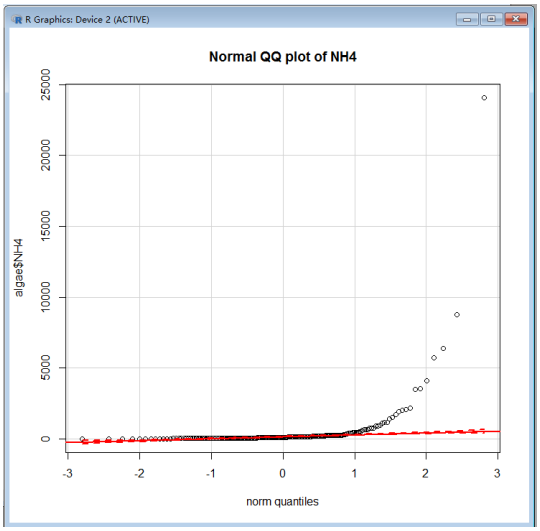
NO3



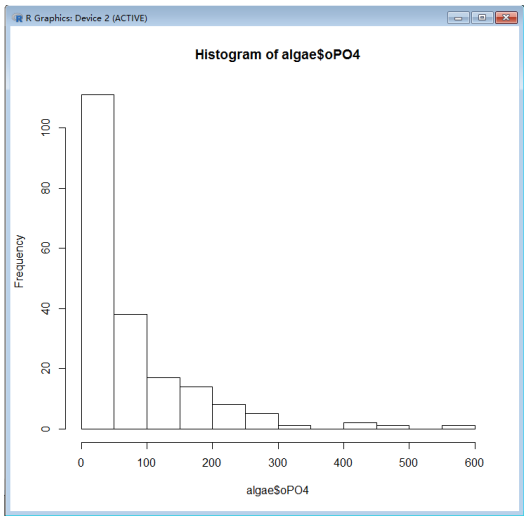


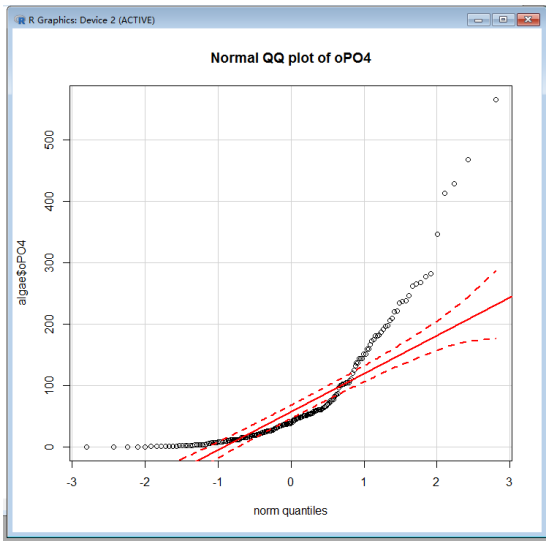
NH4:



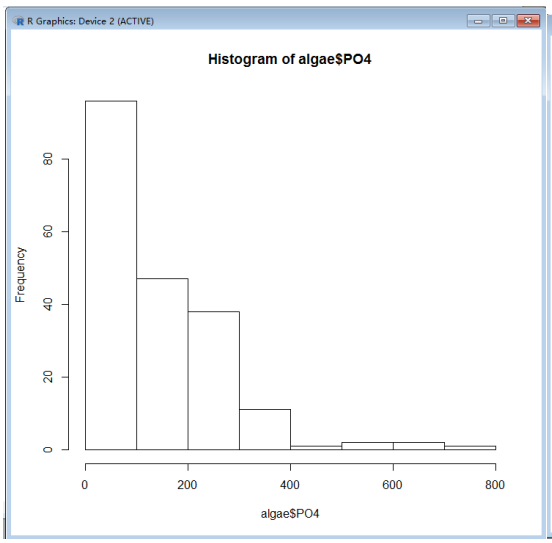


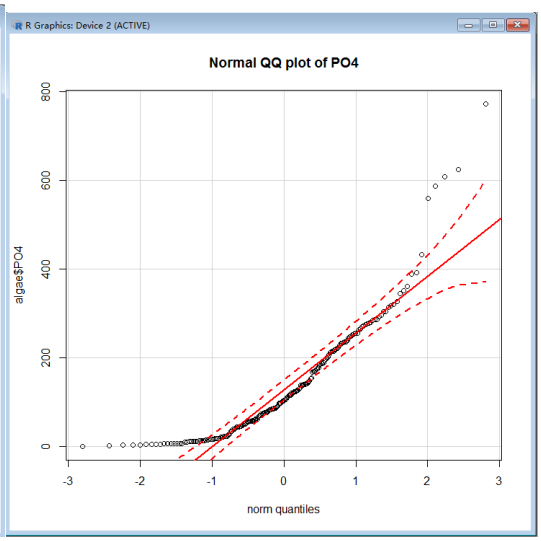
oPO4:



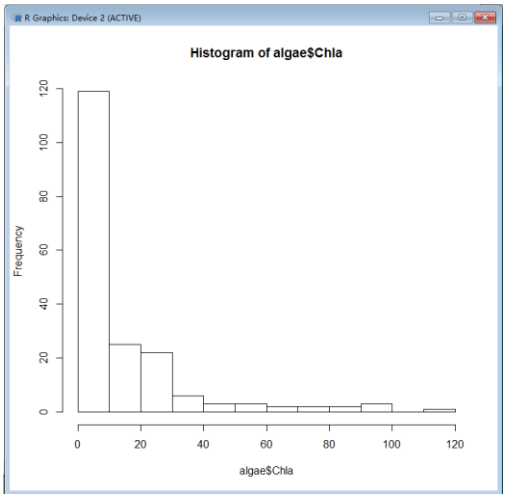


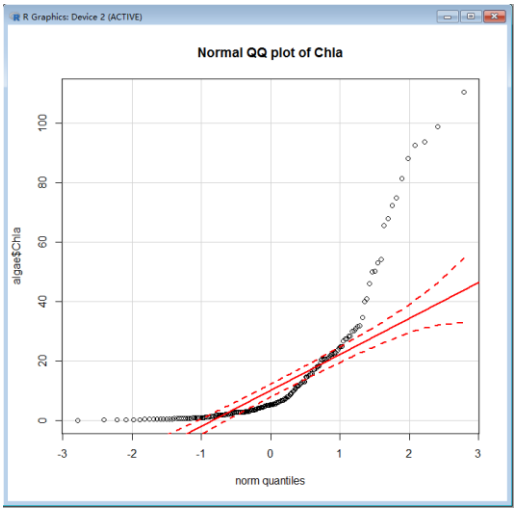
PO4:





Chla:





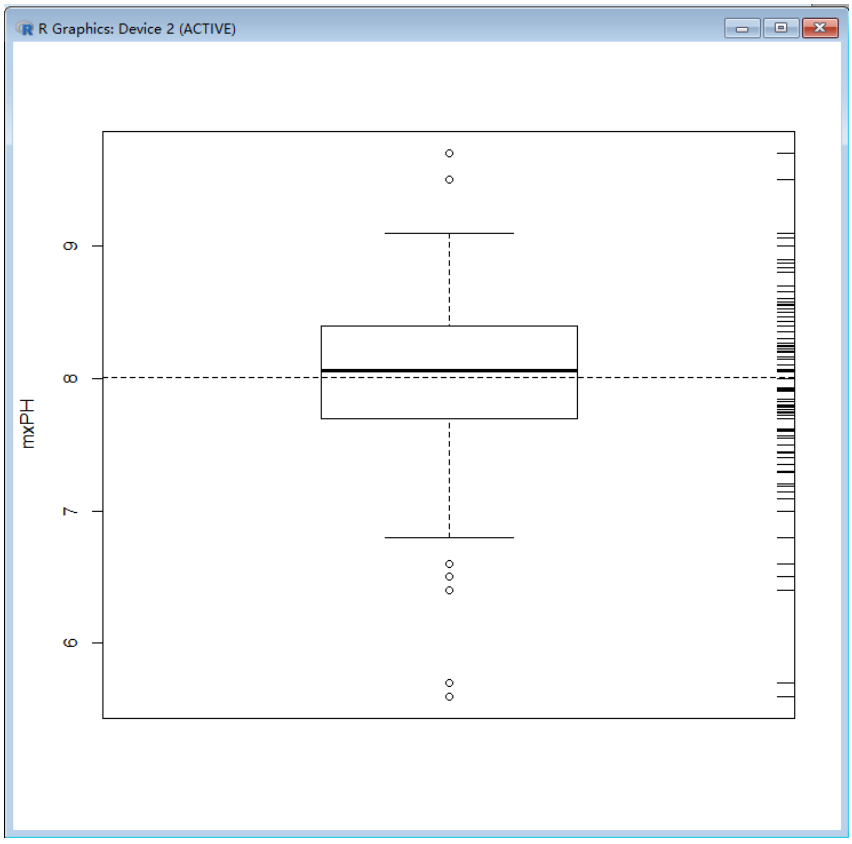
盒图并对离散值进行识别：

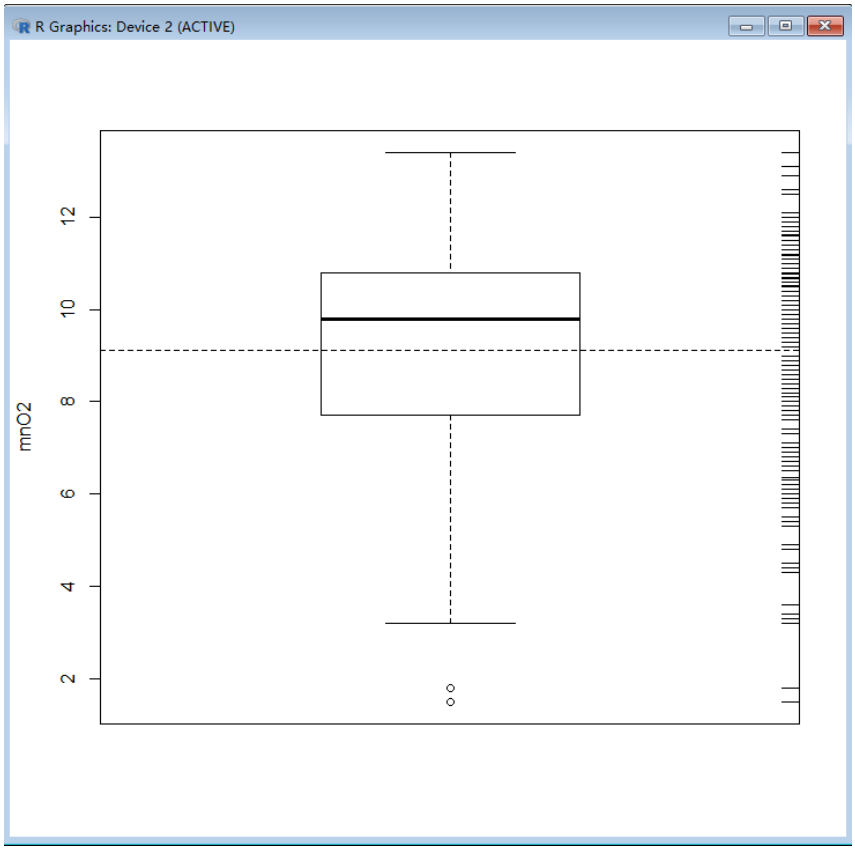
绘制了 oPO4 的盒图，命令行如下

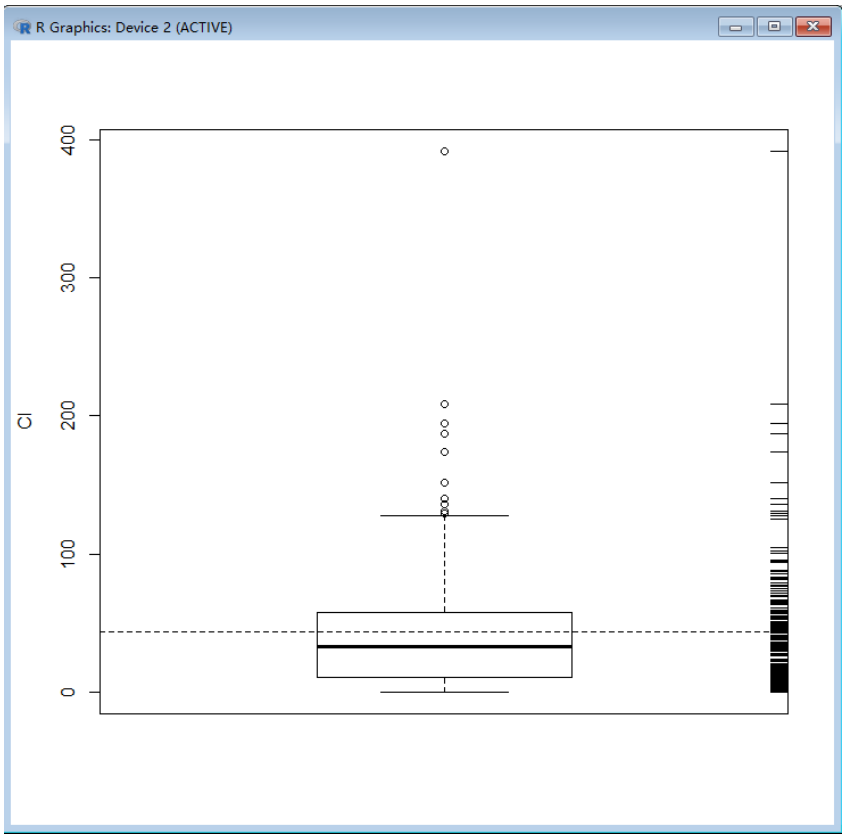
```
boxplot(algae$mxPH,ylab='mxPH')
rug(algae$mxPH,side=4)
abline(h=mean(algae$mxPH,na.rm=T),lty=2)

boxplot(algae$mnO2,ylab='mnO2')
rug(algae$mnO2,side=4)
abline(h=mean(algae$mnO2,na.rm=T),lty=2)

boxplot(algae$Cl,ylab='Cl')
rug(algae$Cl,side=4)
abline(h=mean(algae$Cl,na.rm=T),lty=2)

boxplot(algae$NO3,ylab='NO3')
rug(algae$NO3,side=4)
```

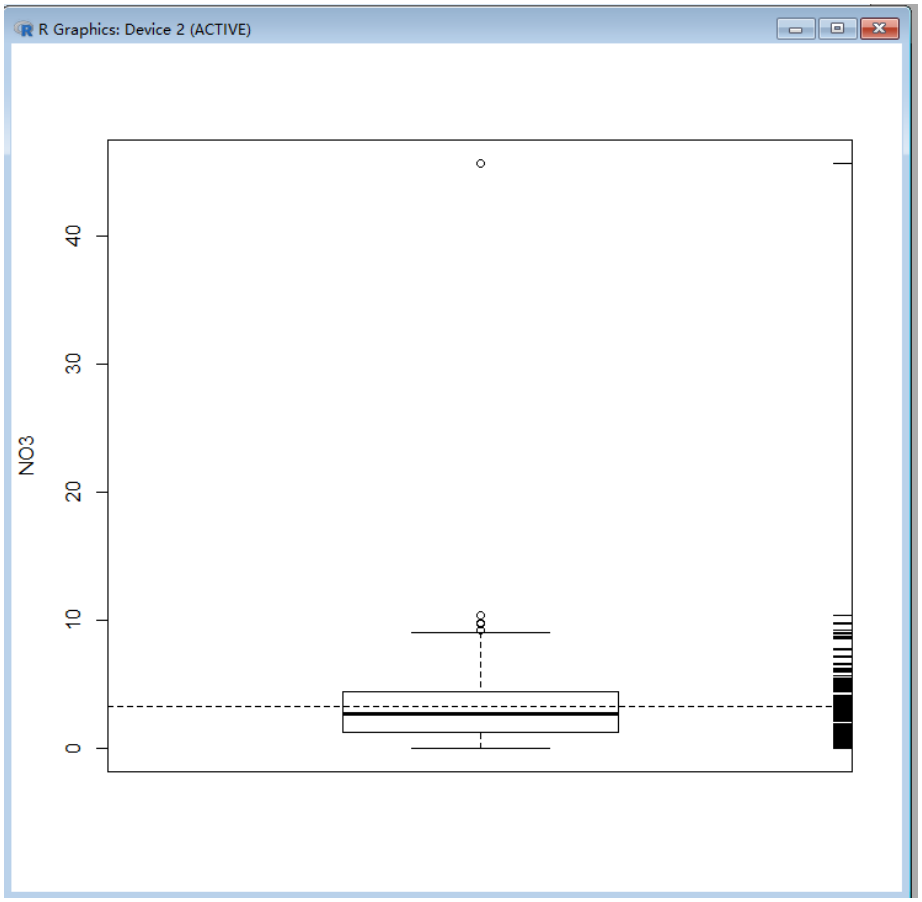
```
abline(h=mean(algae$NO3,na.rm=T),lty=2)

boxplot(algae$NH4,ylab='NH4')
rug(algae$NH4,side=4)
abline(h=mean(algae$NH4,na.rm=T),lty=2)

boxplot(algae$oPO4,ylab='oPO4')
rug(algae$oPO4,side=4)
abline(h=mean(algae$oPO4,na.rm=T),lty=2)

boxplot(algae$PO4,ylab='PO4')
rug(algae$PO4,side=4)
abline(h=mean(algae$PO4,na.rm=T),lty=2)

boxplot(algae$Chla,ylab='Chla')
rug(algae$Chla,side=4)
abline(h=mean(algae$Chla,na.rm=T),lty=2)
```

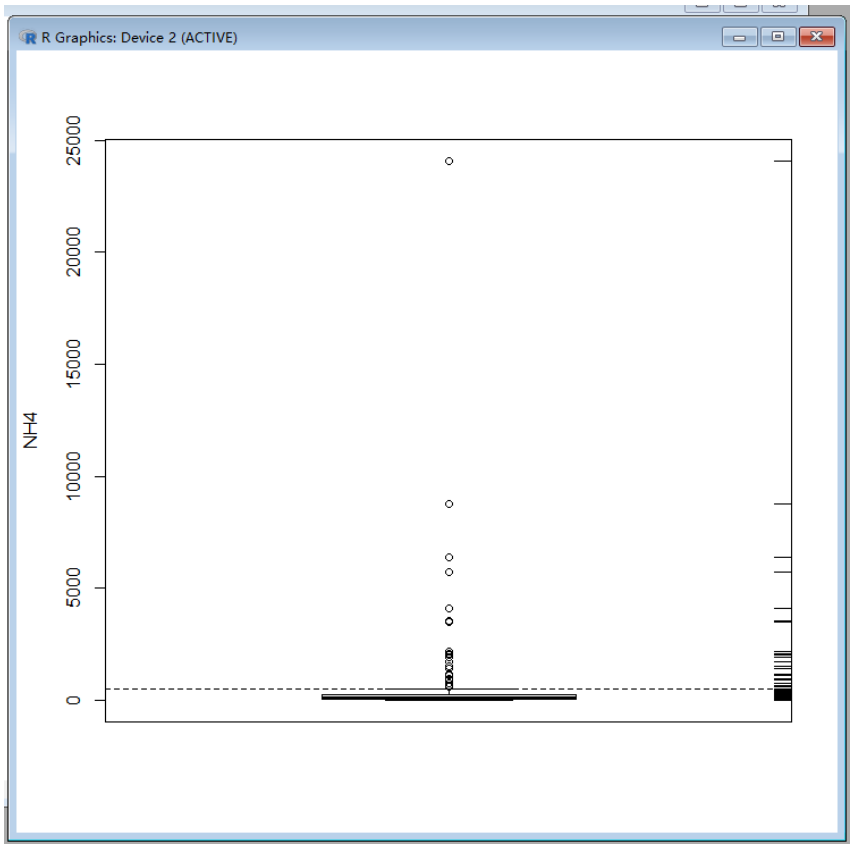
结果如下：

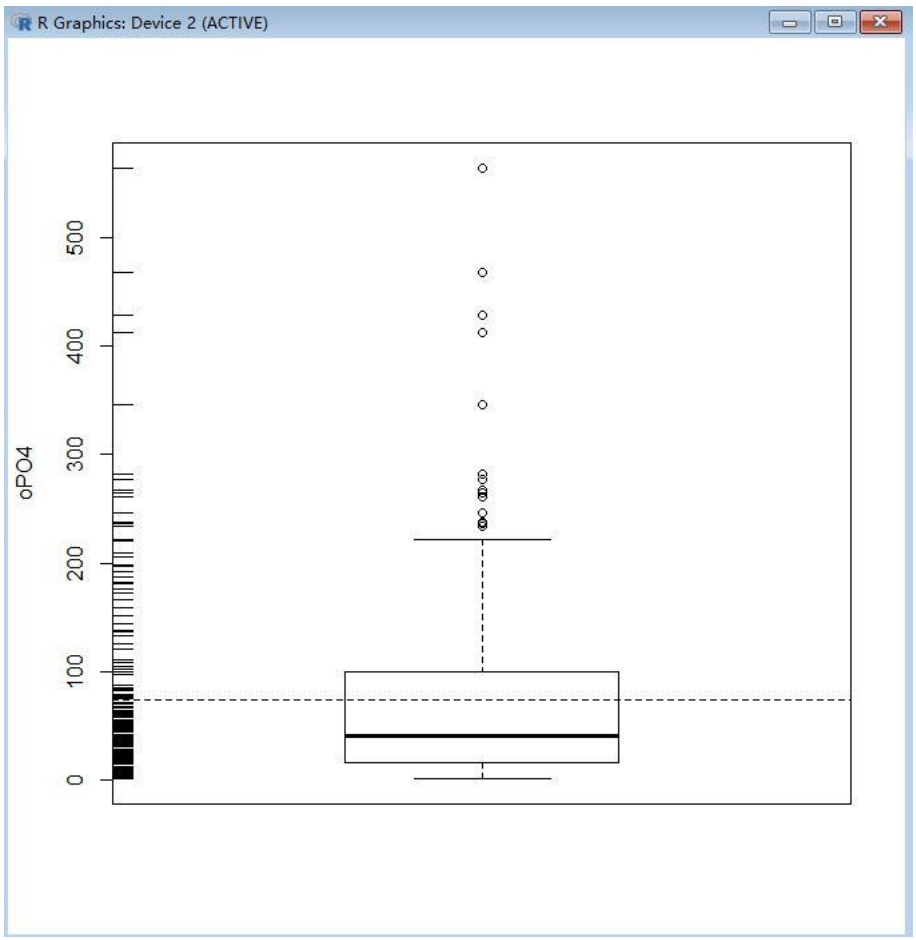


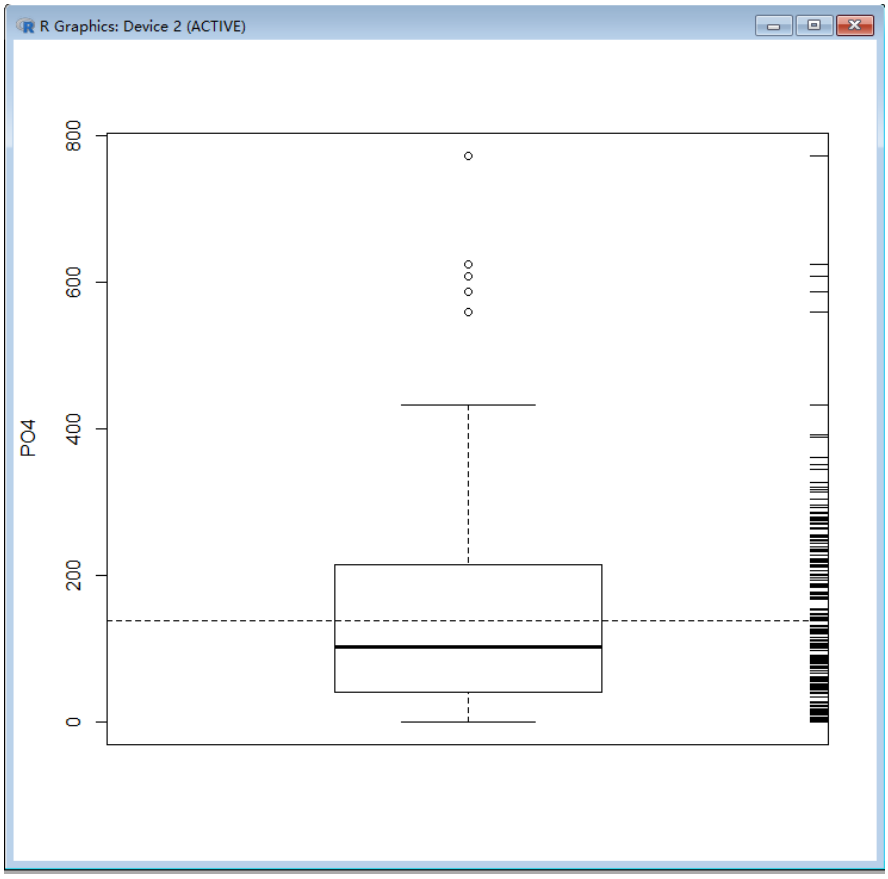

R Graphics: Device 2 (ACTIVE)
mnO2
2
4
6
8
10
12

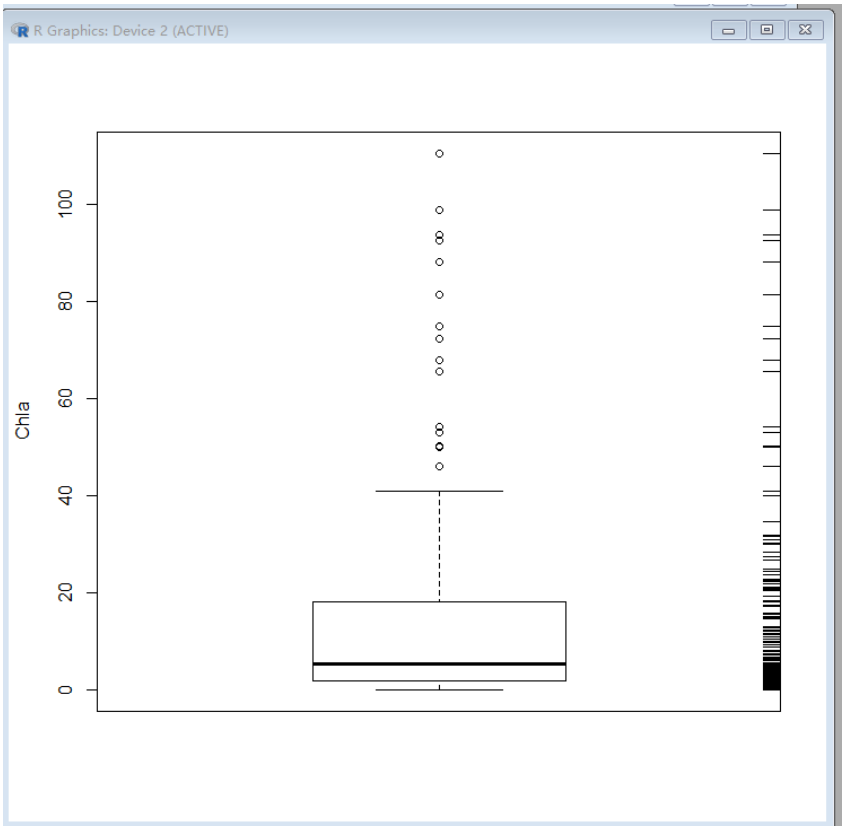

R Graphics: Device 2 (ACTIVE)
Cl
0
100
200
300
400


R Graphics: Device 2 (ACTIVE)
NO3
0
10
20
30
40


R Graphics: Device 2 (ACTIVE)
NH4
0
5000
10000
15000
20000
25000


R Graphics: Device 2 (ACTIVE)
oPO4
0
100
200
300
400
500


R Graphics: Device 2 (ACTIVE)
PO4
0
200
400
600
800



结论：横线上方和下方的圆点表示与其他值相比特比大的值，通常认为是离散值。

条件盒图

条件绘图是依赖于某个特定因子的图形表示，因子是一个为一个取值为有限集合的名义变量。例如，对于 size 的不同取值，可以绘制变量 a1 的一组箱图。每个箱图是对应于变量 size 的某个特定值的水样子集。通过这些箱线图可以研究名义变量 size 如何影响变量 a1 值得分布。

对七种海藻，绘制条件盒图

```
> library(lattice)
> bwplot(size~a1,data=algae,ylab='river size',xlab='Alga A1')
> bwplot(size~a2,data=algae,ylab='river size',xlab='Alga A2')
> bwplot(size~a3,data=algae,ylab='river size',xlab='Alga A3')
> bwplot(size~a4,data=algae,ylab='river size',xlab='Alga A4')
> bwplot(size~a5,data=algae,ylab='river size',xlab='Alga A5')
> bwplot(size~a6,data=algae,ylab='river size',xlab='Alga A6')
> bwplot(size~a7,data=algae,ylab='river size',xlab='Alga A7')
```

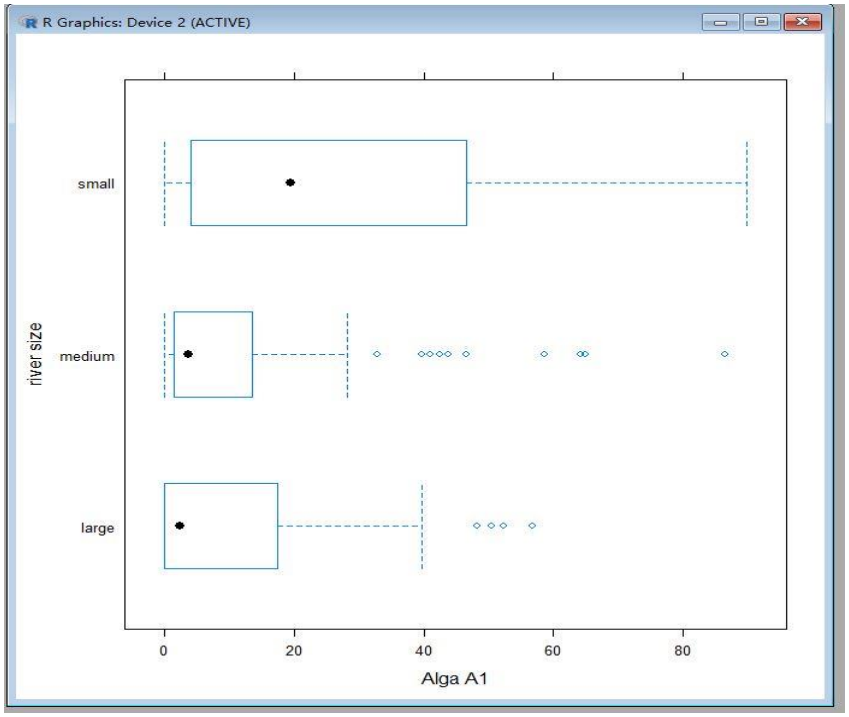
对七种海藻，绘制分位盒图

```
>library(Hmisc)
>
bwplot(size~a1,data=algae,panel=panel.bpplot,probs=seq(.01,.49,by=.01),datadensity=T,ylab='
river size',xlab='Alga A1')
>
bwplot(size~a2,data=algae,panel=panel.bpplot,probs=seq(.01,.49,by=.01),datadensity=T,ylab='
```

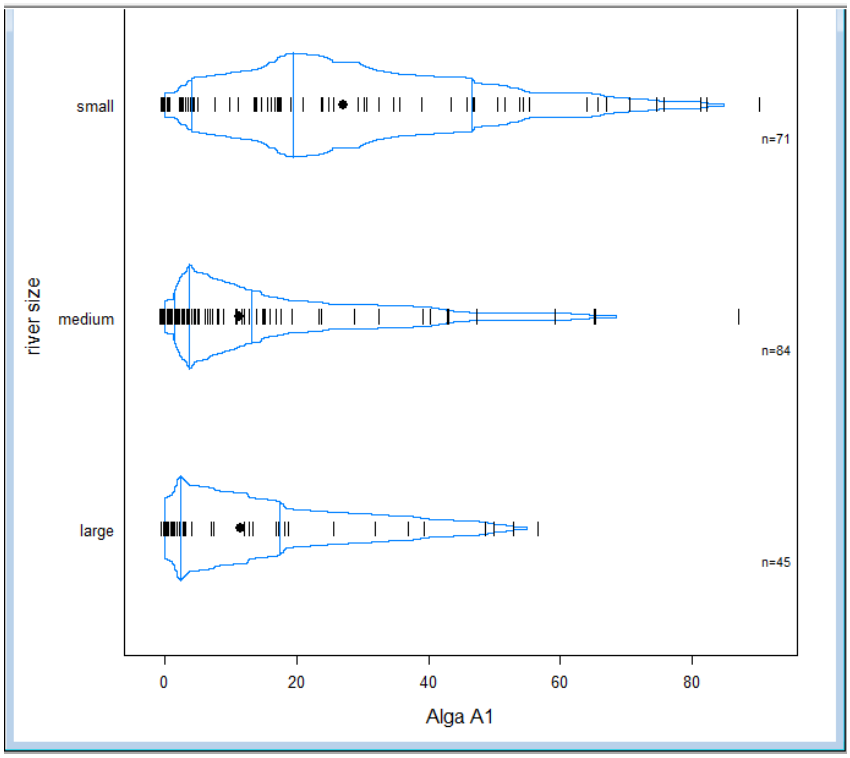
```
river size',xlab='Alga A2')
>
bwplot(size~a3,data=algae,panel=panel.bpplot,probs=seq(.01,.49,by=.01),datadensity=T,ylab='
river size',xlab='Alga A3')
>
bwplot(size~a4,data=algae,panel=panel.bpplot,probs=seq(.01,.49,by=.01),datadensity=T,ylab='
river size',xlab='Alga A4')
>
bwplot(size~a5,data=algae,panel=panel.bpplot,probs=seq(.01,.49,by=.01),datadensity=T,ylab='
river size',xlab='Alga A5')
>
bwplot(size~a6,data=algae,panel=panel.bpplot,probs=seq(.01,.49,by=.01),datadensity=T,ylab='
river size',xlab='Alga A6')
>
bwplot(size~a7,data=algae,panel=panel.bpplot,probs=seq(.01,.49,by=.01),datadensity=T,ylab='
river size',xlab='Alga A7')
```

结果如下,上面为条件盒图，下面为分位箱图。分位箱图中的点代表代表不同大小的河流中海藻频数的均值，而图中的竖线分别代表变量的第一份位数，中位数和第三分位数。图中的小竖线代表数据的真实取值，这些值分布信息由分位数图来体现。分位数箱图提供的信息要多于传统的箱图。例如我们可以确认上面的观测结论，小型河流有更高频率的海藻，但我们也观察到小型河流的海藻频率的分布比其他类型的河流的海藻频率的分布分散。

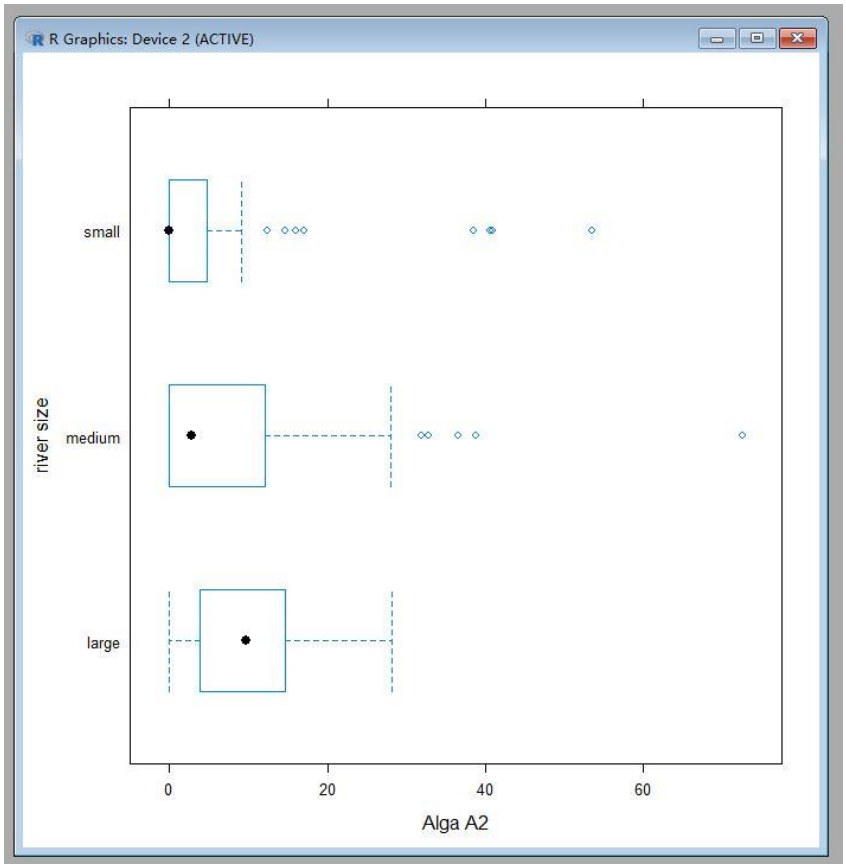
a1:条件盒图：


R Graphics: Device 2 (ACTIVE)
small
medium
large
river size
0
20
40
60
80
Alga A1

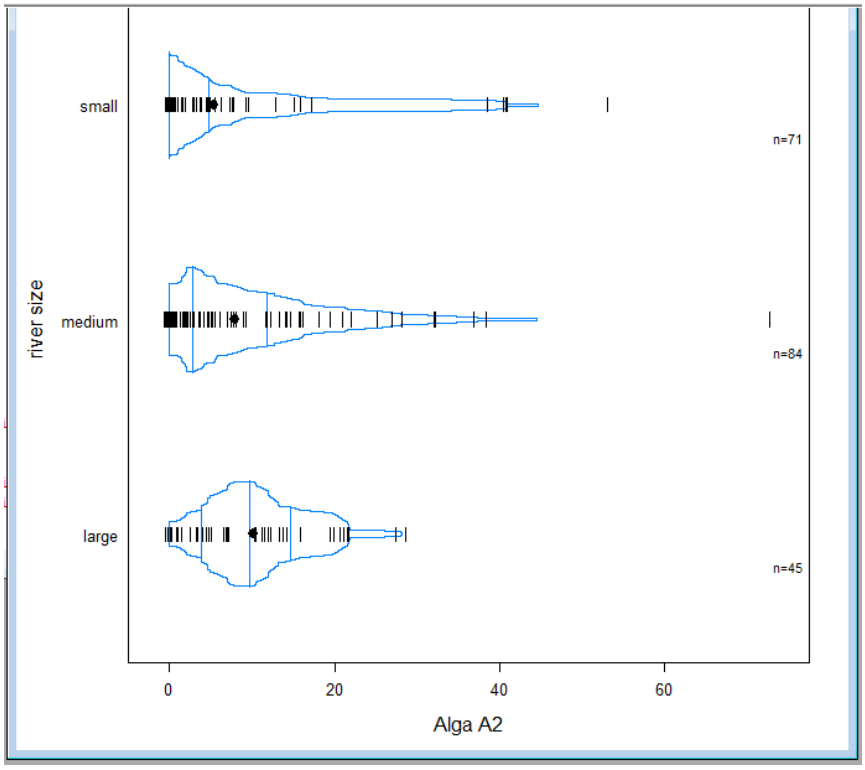
分位箱图


small
medium
large
river size
n=71
n=84
n=45
0
20
40
60
80
Alga A1

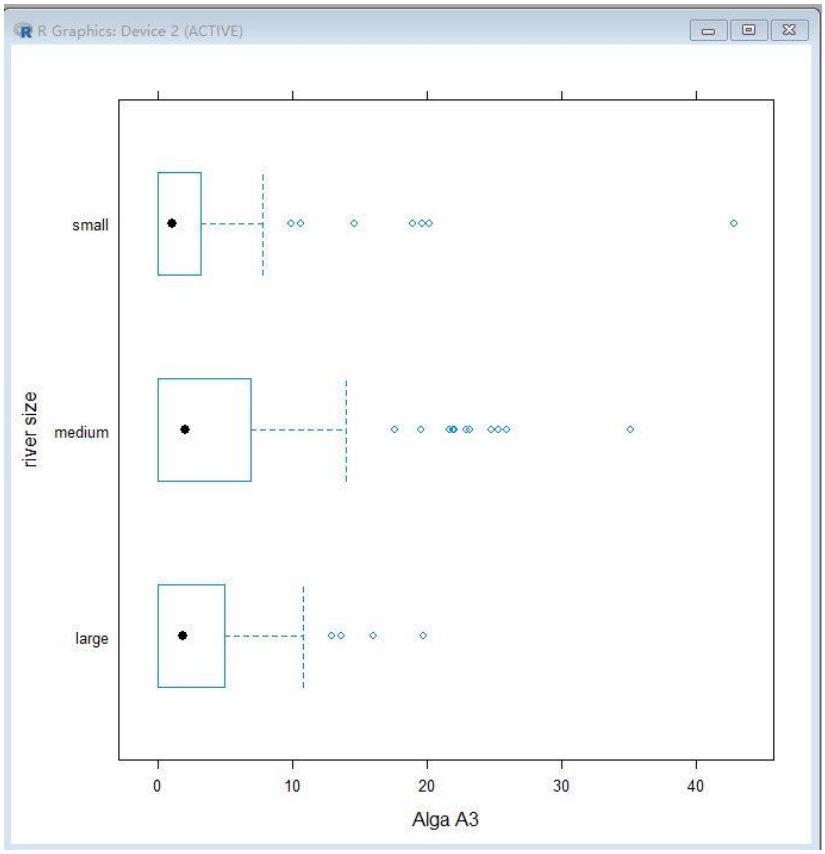
a2 条件盒图:


R Graphics: Device 2 (ACTIVE)
small
medium
large
river size
0
20
40
60
Alga A2

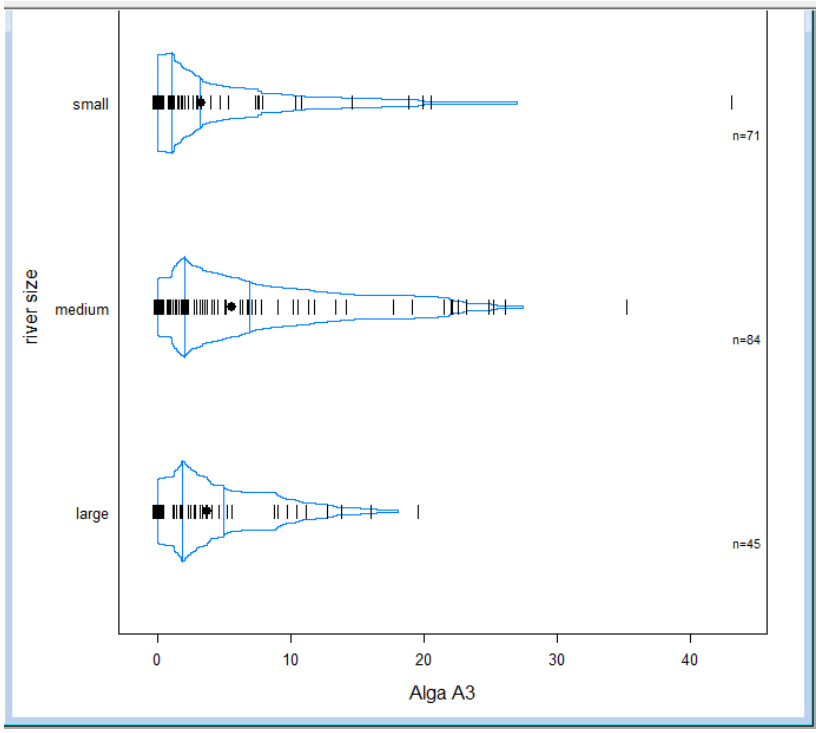
分位箱图：


small
medium
large
river size
n=71
n=84
n=45
0
20
40
60
Alga A2

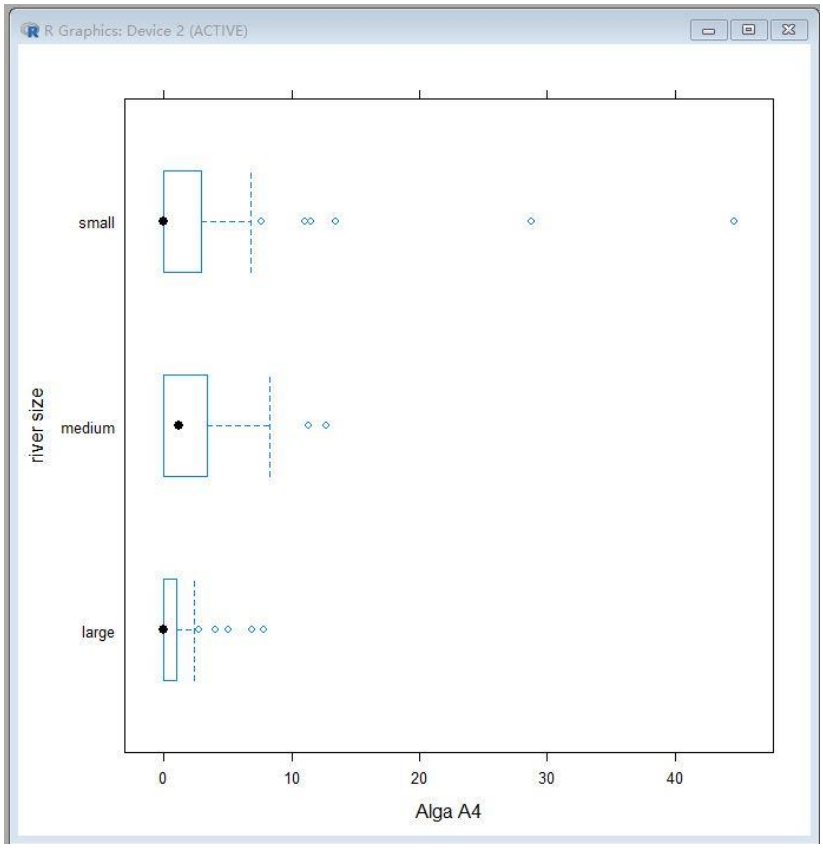
a3:条件盒图


R Graphics: Device 2 (ACTIVE)
small
medium
large
river size
0
10
20
30
40
Alga A3

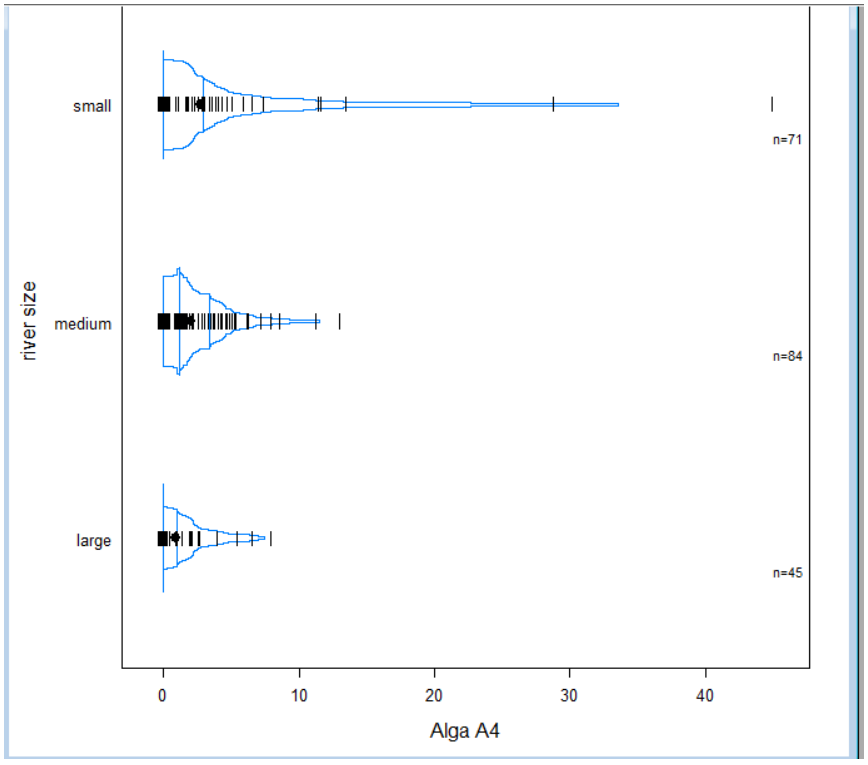
分位箱图：


small
medium
large
river size
n=71
n=84
n=45
0
10
20
30
40
Alga A3

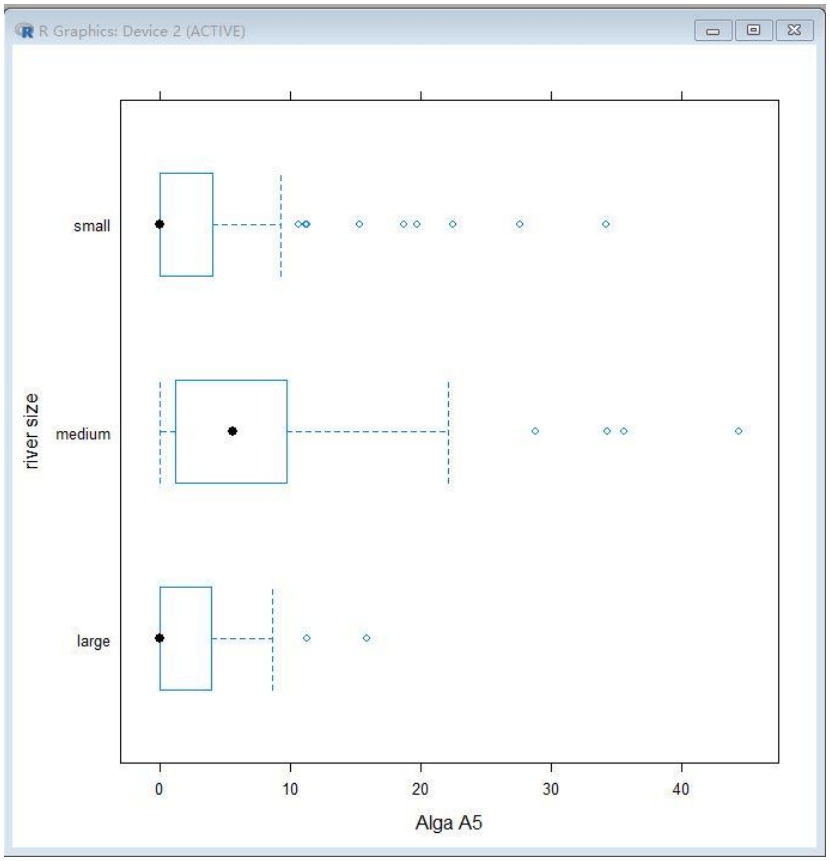
a4:条件盒图


R Graphics: Device 2 (ACTIVE)
small
medium
large
river size
0
10
20
30
40
Alga A4

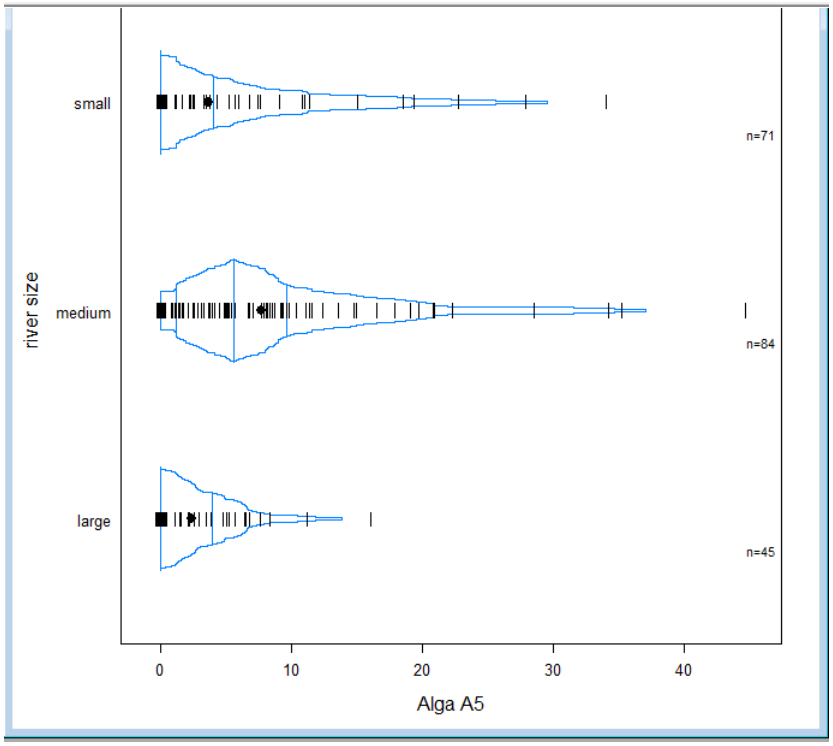
分位箱图


small
medium
large
river size
n=71
n=84
n=45
0
10
20
30
40
Alga A4

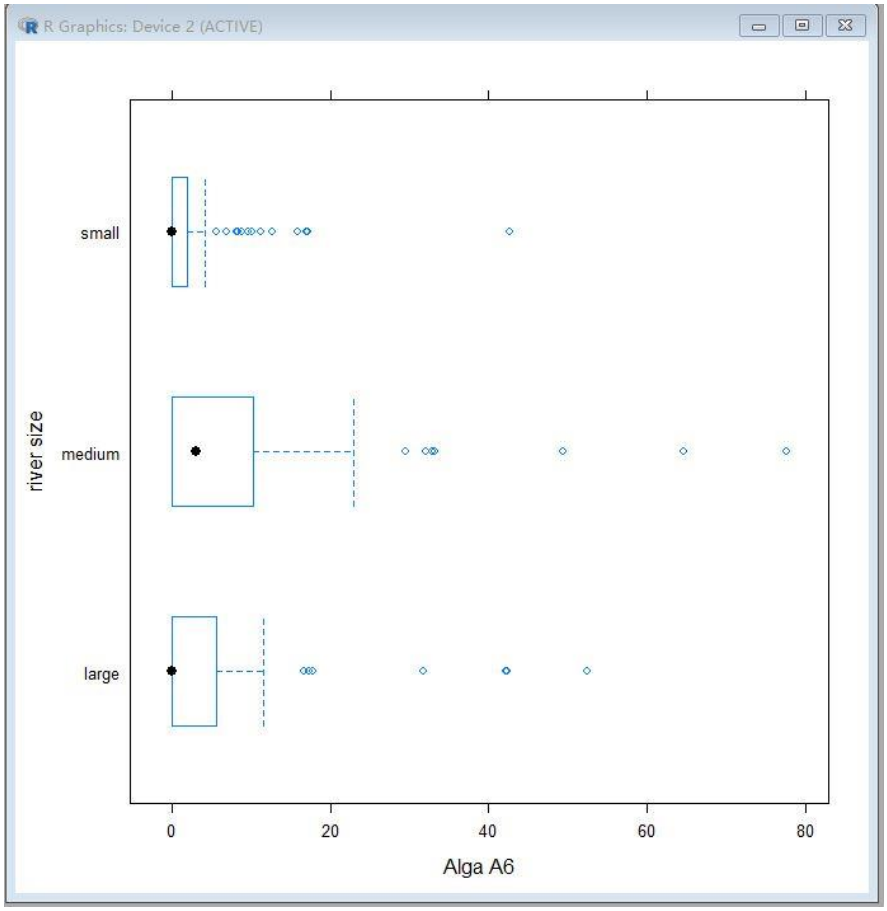
a5:条件盒图


R Graphics: Device 2 (ACTIVE)
small
medium
large
river size
0
10
20
30
40
Alga A5

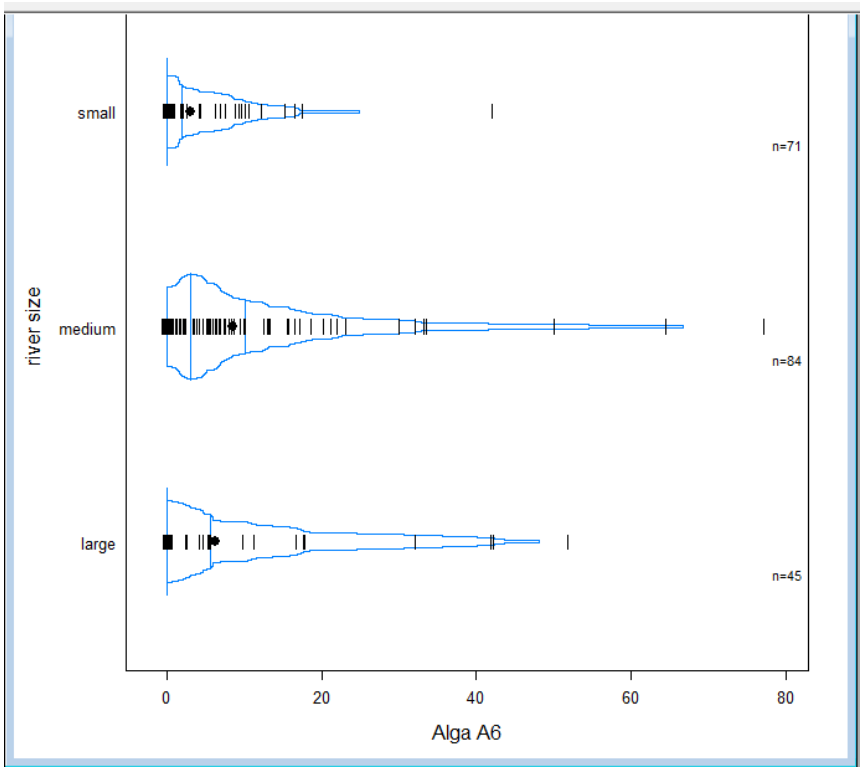
分位箱图


small
medium
large
river size
n=71
n=84
n=45
0
10
20
30
40
Alga A5

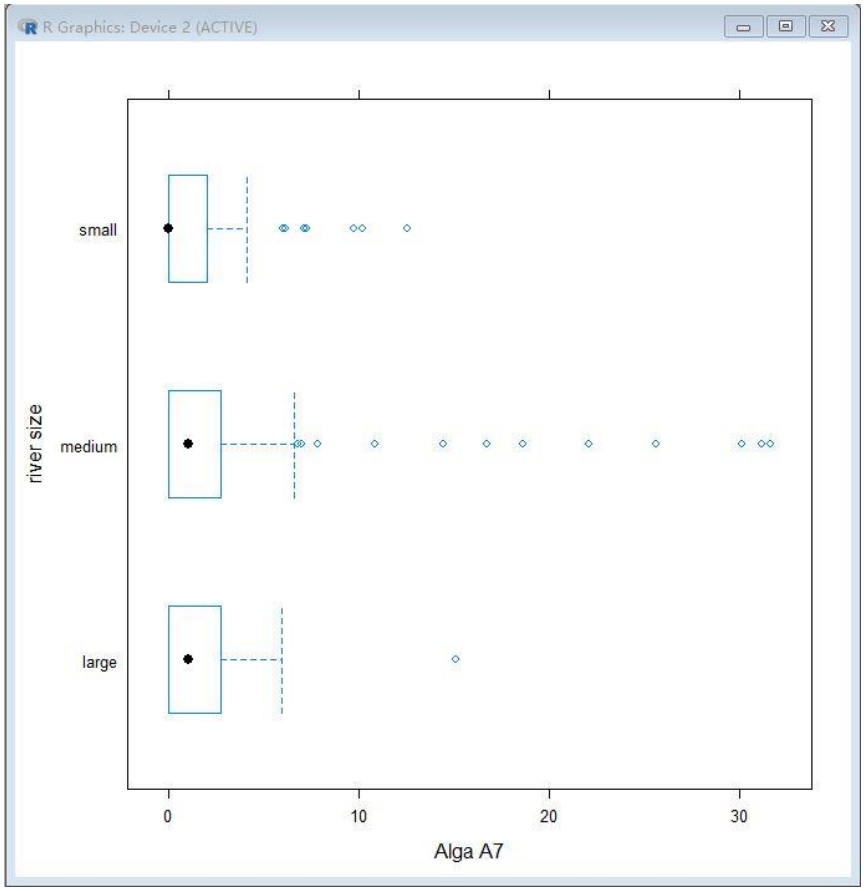
a6:条件盒图


R Graphics: Device 2 (ACTIVE)
small
medium
large
river size
0
20
40
60
80
Alga A6

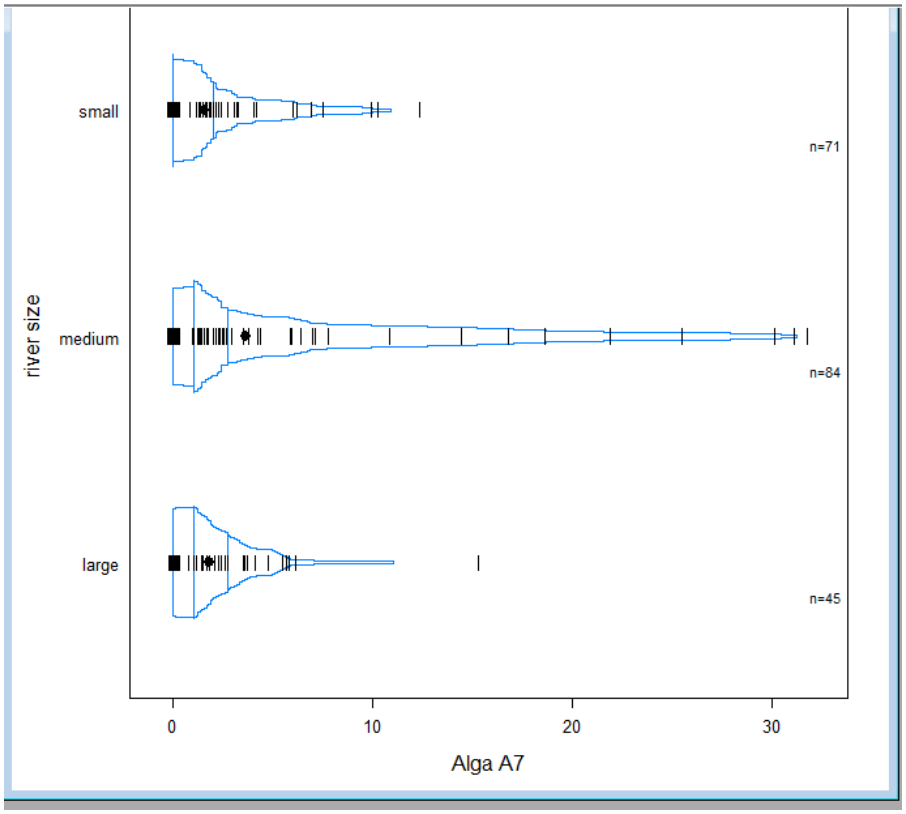
分位箱图


small
medium
large
river size
n=71
n=84
n=45
0
20
40
60
80
Alga A6

a7:条件盒图


R Graphics: Device 2 (ACTIVE)
small
medium
large
river size
0
10
20
30
Alga A7

分位箱图：


small
medium
large
river size
n=71
n=84
n=45
0
10
20
30
Alga A7

数据缺失的处理

将缺失部分剔除

首先检查含缺失值的记录，然后剔除所含缺失值的记录。命令行如下：

```
> algae[!complete.cases(algae),]
> nrow(algae[!complete.cases(algae),])
> algae<-na.omit(algae)
```

结果：

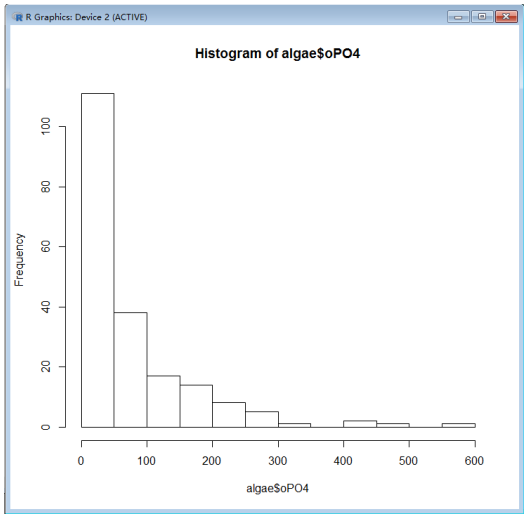
```
> algae[!complete.cases(algae),]
    season   size  speed mxPH mnO2    Cl   NO3 NH4    oPO4     PO4  Chla   a1   a2  a3   a4  a5  a6  a7
28  autumn  small   high 6.80 11.1 9.000 0.630  20   4.000      NA  2.70 30.3  1.9 0.0  0.0 2.1 1.4 2.1
38  spring  small   high 8.00   NA 1.450 0.810  10   2.500   3.000  0.30 75.8  0.0 0.0  0.0 0.0 0.0 0.0
48  winter  small    low   NA 12.6 9.000 0.230  10   5.000   6.000  1.10 35.5  0.0 0.0  0.0 0.0 0.0 0.0
55  winter  small   high 6.60 10.8    NA 3.245  10   1.000   6.500    NA 24.3  0.0 0.0  0.0 0.0 0.0 0.0
56  spring  small medium 5.60 11.8    NA 2.220   5   1.000   1.000    NA 82.7  0.0 0.0  0.0 0.0 0.0 0.0
57  autumn  small medium 5.70 10.8    NA 2.550  10   1.000   4.000    NA 16.8  4.6 3.9 11.5 0.0 0.0 0.0
58  spring  small   high 6.60  9.5    NA 1.320  20   1.000   6.000    NA 46.8  0.0 0.0 28.8 0.0 0.0 0.0
59  summer  small   high 6.60 10.8    NA 2.640  10   2.000  11.000    NA 46.9  0.0 0.0 13.4 0.0 0.0 0.0
60  autumn  small medium 6.60 11.3    NA 4.170  10   1.000   6.000    NA 47.1  0.0 0.0  0.0 0.0 1.2 0.0
61  spring  small medium 6.50 10.4    NA 5.970  10   2.000  14.000    NA 66.9  0.0 0.0  0.0 0.0 0.0 0.0
62  summer  small medium 6.40   NA    NA    NA  NA      NA  14.000    NA 19.4  0.0 0.0  2.0 0.0 3.9 1.7
63  autumn  small   high 7.83 11.7 4.083 1.328  18   3.333   6.667    NA 14.4  0.0 0.0  0.0 0.0 0.0 0.0
116 winter medium   high 9.70 10.8 0.222 0.406  10  22.444  10.111    NA 41.0  1.5 0.0  0.0 0.0 0.0 0.0
161 spring  large    low 9.00  5.8    NA 0.900 142 102.000 186.000 68.05  1.7 20.6 1.5  2.2 0.0 0.0 0.0
184 winter  large   high 8.00 10.9 9.055 0.825  40  21.083  56.091    NA 16.8 19.6 4.0  0.0 0.0 0.0 0.0
199 winter  large medium 8.00  7.6    NA    NA  NA      NA      NA    NA  0.0 12.5 3.7  1.0 0.0 0.0 4.9
> nrow(algae[!complete.cases(algae),])
[1] 16
```

上图给出了缺失的记录，并且给出缺失值的记录条数为 16

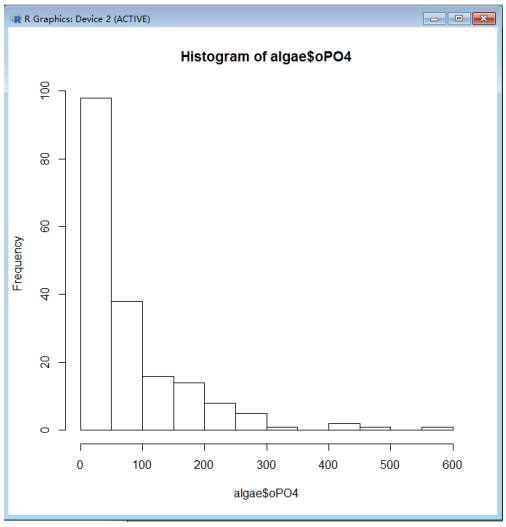
然后进行剔除，剔除后缺失值为 0

```
> algae<-na.omit(algae)
> nrow(algae[!complete.cases(algae),])
[1] 0
```

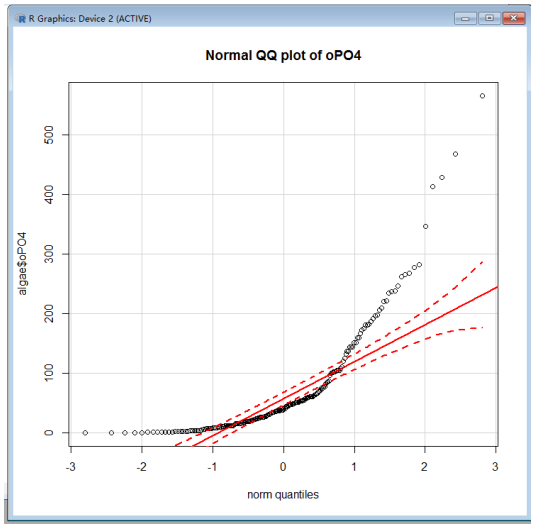
对比结果，此处选取一个属性来进行对比，选取 oPO4 属性，结果如下：



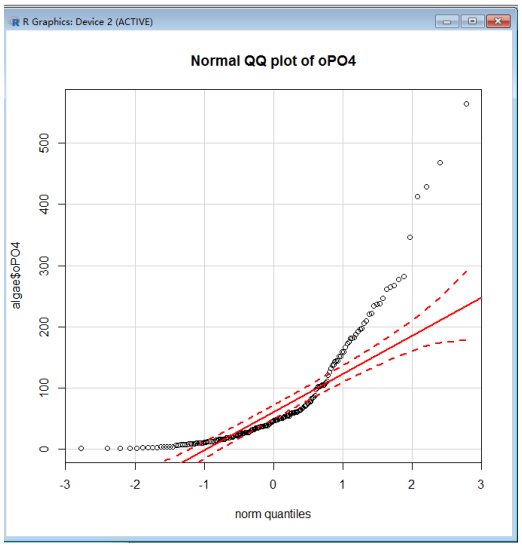
1，oPO4 处理前的直方图



2, oPO4 经过删除缺失值处理后的直方图



3 oPO4 处理前的 QQ 图



4　oPO4 经过删除缺失值处理后的 QQ 图

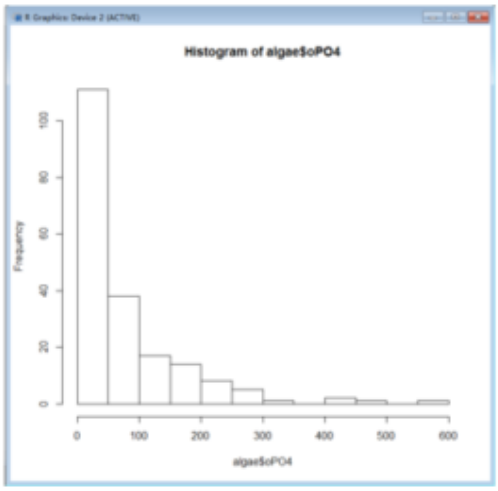
用最高频率值来填补缺失值

填补缺失数据最简单和快捷的方法是使用一些代表中心趋势的值，代表中心趋势的值反映了变量分布的最常见的值，因此中心趋势值是最自然的选择。有多个代表数据中心趋势的指标，例如平均值，中位数，众数等。最合适的选择由变量的分布决定。对于接近正态的分布来说，所有的观测值都较好地聚集在平均值周围，平均值数就是最佳选择。然而，对于偏态分布，或者离群值的变量来说，选择平均值就不好。偏态分布的大部分值都聚集在变量分布的一侧，因此平均值不能作为最常见的代表。另一方面，离群值（极值）的存在会扭曲平均值，这就导致了平均值不具有代表性的问题。因此，在对变量分布进行检查之前选择平均值作为中心趋势的代表是不明智的，例如某些 R 的绘制工具。对偏态分布或者有离群值的分布而言，中位数是更好的代表数据中心趋势的指标。

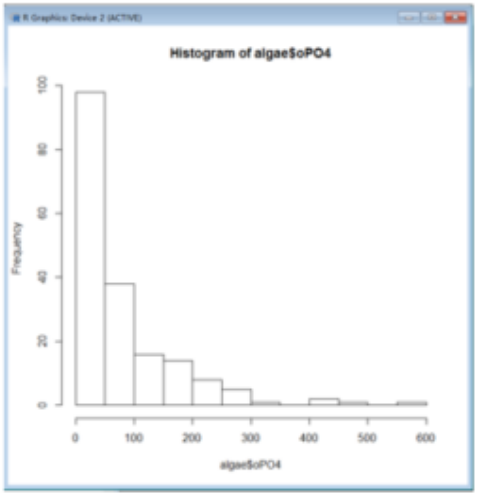
代码

```
> algae<-read.table('C:/HZTAO/course/dataMining/R/Analysis.txt',header=T,dec='.',na.strings=c('XXXXXXX'))
> nrow(algae[!complete.cases(algae),])
[1] 16
> algae[is.na(algae$season),'season']<-"winter"
> algae[is.na(algae$size),'size']<-"medium"
> algae[is.na(algae$speed),'speed']<-"high"
> algae[is.na(algae$mxPH),'mxPH']<-median(algae$mxPH,na.rm=T)
> algae[is.na(algae$mnO2),'mnO2']<-median(algae$mnO2,na.rm=T)
> algae[is.na(algae$Cl),'Cl']<-median(algae$Cl,na.rm=T)
> algae[is.na(algae$NO3),'NO3']<-median(algae$NO3,na.rm=T)
> algae[is.na(algae$NH4),'NH4']<-median(algae$NH4,na.rm=T)
> algae[is.na(algae$oPO4),'oPO4']<-median(algae$oPO4,na.rm=T)
> algae[is.na(algae$PO4),'PO4']<-median(algae$PO4,na.rm=T)
> algae[is.na(algae$Chla),'Chla']<-median(algae$Chla,na.rm=T)
> algae[is.na(algae$a1),'a1']<-median(algae$a1,na.rm=T)
> algae[is.na(algae$a2),'a2']<-median(algae$a2,na.rm=T)
> algae[is.na(algae$a3),'a3']<-median(algae$a3,na.rm=T)
> algae[is.na(algae$a4),'a4']<-median(algae$a4,na.rm=T)
> algae[is.na(algae$a5),'a5']<-median(algae$a5,na.rm=T)
> algae[is.na(algae$a6),'a6']<-median(algae$a6,na.rm=T)
> algae[is.na(algae$a7),'a7']<-median(algae$a7,na.rm=T)
```

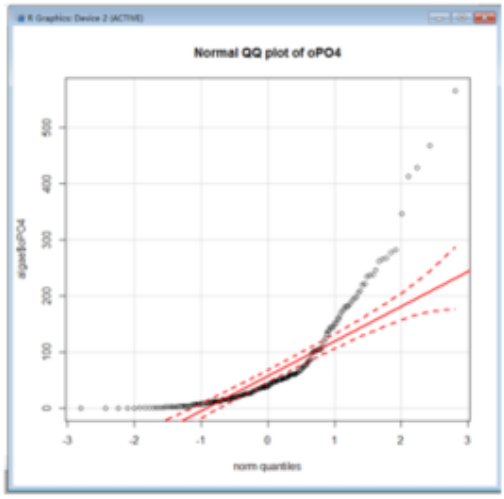
对比结果：还是选取属性 oPO4 来进行对比。



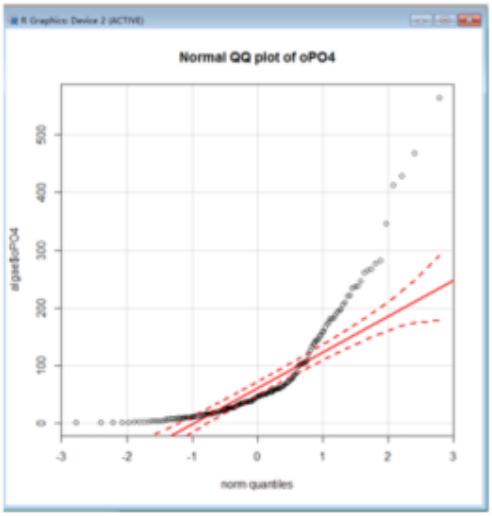
1，oPO4 处理前的直方图



2，oPO4 经过高频率处理后的直方图



3 oPO4 处理前的 QQ 图



4 oPO4 经过高频率处理后的 QQ 图

通过变量的相关关系来调补缺失值

一种获取缺失值较少偏差估计值的方法是探寻变量之间的相关关系。
过程如下：
1：获取变量之间的相关矩阵：

```
[1] 0
> algae<-read.table('C:/HZTAO/course/dataMining/R/Analysis.txt',hea
> nrow(algae[!complete.cases(algae),])
[1] 16
> symnum(cor(algae[,4:18],use="complete.obs"))
     mP mO Cl NO NH o P Ch a1 a2 a3 a4 a5 a6 a7
mxPH 1
mnO2    1
Cl         1
NO3           1
NH4           ,  1
oPO4    .  .        1
PO4     .  .        * 1
Chla .                   1
a1         .        . .    1
a2   .                   .     1
a3                              1
a4      .           . .             1
a5                                     1
a6           .  .                      .  1
a7                                            1
attr(,"legend")
[1] 0 ' ' 0.3 '.' 0.6 ',' 0.8 '+' 0.9 '*' 0.95 'B' 1
>
```

Cor（）函数产生变量之间的相关值矩阵。
2：结果显示，NH4 和 NO3 之间，PO4 和 oPO4 之间相关性较大。在 NH4 和 NO3 之间，相关性不是特别明显，而且只有两条样本含有过多缺失值，剔除它们则不存在缺失值。PO4 和 oPO4 的相关性很大，可以用变量的相关性填补缺失值。
3：寻找 PO4 和 oPO4 之间的线性关系：

> lm(formula=PO4~oPO4,data=algae)

```
> lm(formula=PO4~oPO4,data=algae)

Call:
lm(formula = PO4 ~ oPO4, data = algae)

Coefficients:
(Intercept)         oPO4
     42.897        1.293

>
```

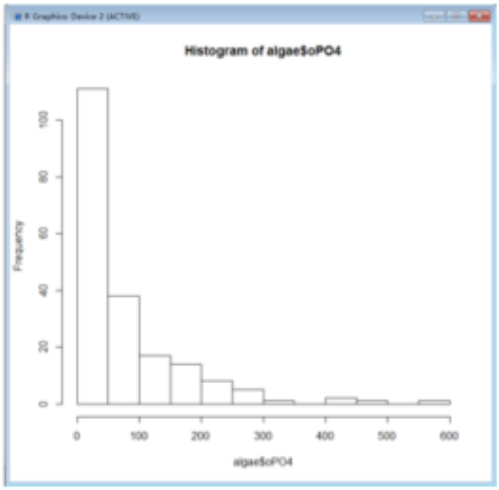
则线性模型为：PO4 = 42.897+1.293*oPO4。
4：NH4 和 NO3 剔除样本 62 和 199 后，则只有样本 28 在 PO4 上有缺失值，可以用上面的线性关系来填补缺失值。
>algae[28,"PO4"]<-42.897+1.293*algae[28,"oPO4"]

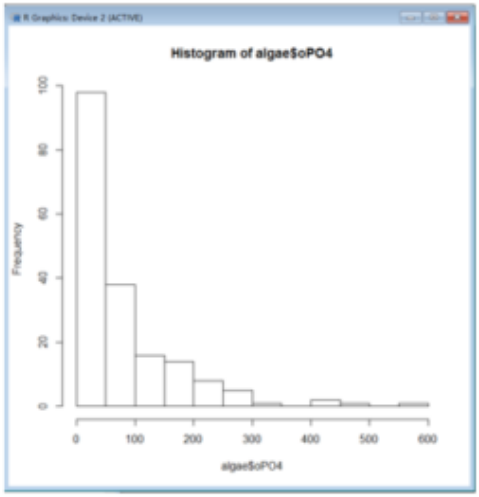
查看插补后的记录，可以看出 PO4 的插补值为 48

> algae[28,]

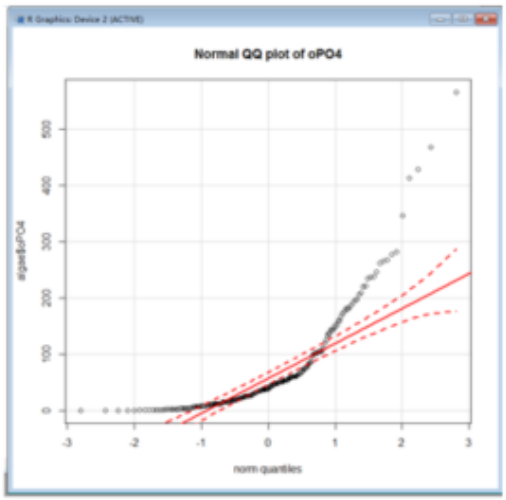
```
> algae[28,"PO4"]<-42.897+1.293*algae[28,"oPO4"]
> algae[28,]
   season  size speed mxPH mnO2 Cl  NO3 NH4 oPO4    PO4 Chla   a1  a2 a3 a4  a5  a6  a7
28 autumn small  high  6.8 11.1  9 0.63  20    4 48.069  2.7 30.3 1.9  0  0 2.1 1.4 2.1
> |
```

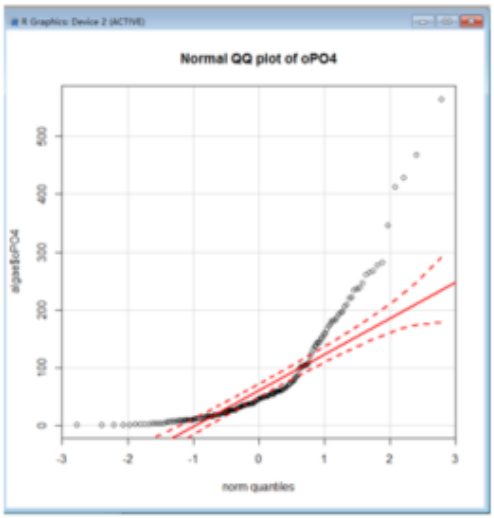
结果对比，选取一个属性值进行对比，如下：





1，oPO4 处理前的直方图　2, oPO4 经过相关关系处理后的直方图



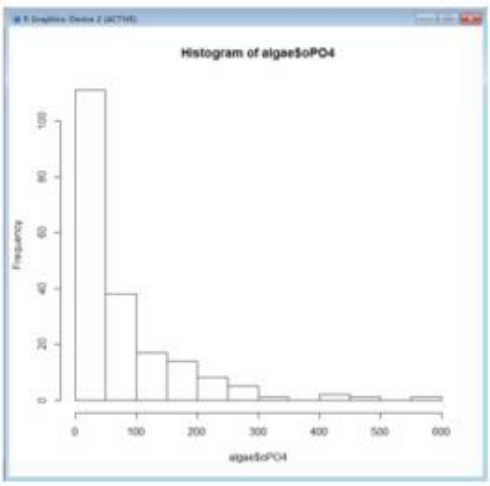


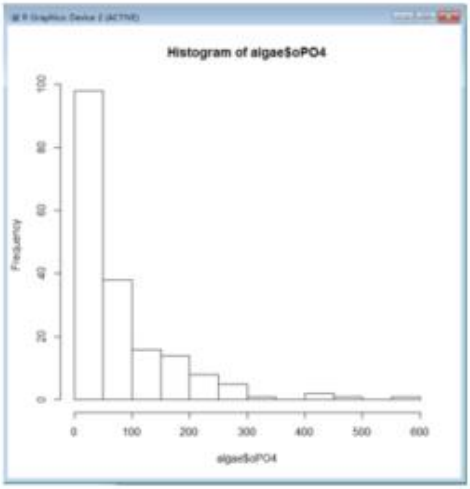
3 oPO4 处理前的 QQ 图　4　oPO4 经过相关关系处理后的 QQ 图

通过数据对象之间的相似性来填补缺失值

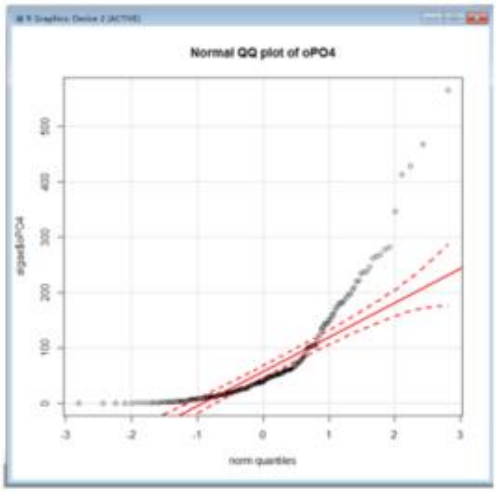
在这里使用欧式距离进行度量相似性。通过这种度量的方法来寻找与任何含有缺失值的数据对象最相似的十个水样，并用它们来填补缺失值。
我们通过添加包函数 knnImputation()来实现，此函数用一个欧式距离的变种来找距离任何个数据最近的 k 个数据。在计算距离时都要对数据进行标准化。
代码：

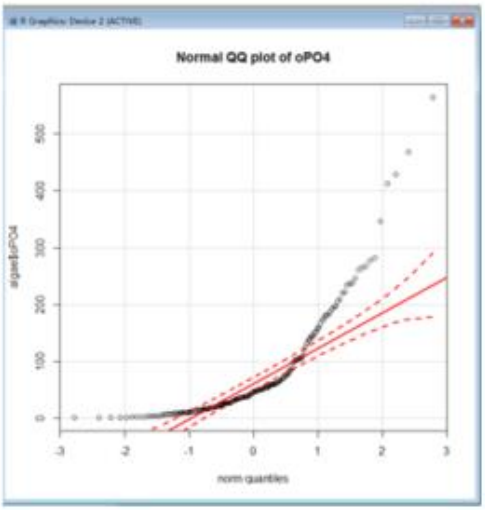
```
clean.algae<-knnImputation(algae,k=10)
```





1，oPO4 处理前的直方图　　2, oPO4 经过相似性处理后的直方图





3 oPO4 处理前的 QQ 图　　4　oPO4 经过相似性处理后的 QQ 图