

Balanced Acceptance Sampling (BAS) for Points

Trent McDonald

July 13, 2016

This vignette demonstrates basic Balanced Acceptance Sampling (BAS) of a point resource. BAS will sample a simple point resource consisting of a 10 X 10 grid of points evenly spaced on a square study area with width 100. The following code creates the example population.

```
library(sp)
# Create the bounding study area as a single polygon
study.area <- Polygon( data.frame(x=c(0,100,100,0,0), y=c(0,0,100,100,0)) )
study.area <- Polygons(list(study.area), "1")
study.area <- SpatialPolygons(list(study.area))
# Create population of points
popln.points <- GridTopology( c(5,5), c(10,10), c(10,10) )
popln.points <- SpatialGrid(popln.points)
popln.points <- SpatialPoints(popln.points)
# Create population of pixels for visualization and sample selection
popln.pixels <- SpatialPixelsDataFrame(popln.points, data.frame(val=1:100))
```

Plot with Halton boxes

The Halton sequence is an integral part of BAS. See Robertson et al. (2015) and documentation of `halton()` for ways to compute the Halton sequence. Here, I plot the Halton boxes induced by using base 2 in the horizontal direction and base 3 in the vertical.

```
plot(popln.pixels, what="image", axes=TRUE, xaxt="n", yaxt="n")
axis(2, at=seq(5,100,by=10), labels = FALSE)
axis(2, at=seq(5,95,by=20), labels = seq(5,95,by=20))
axis(1, at=seq(5,100,by=10), labels = FALSE)
axis(1, at=seq(5,95,by=20), labels = seq(5,95,by=20))
points(popln.points, pch=16, col="red")

abline(v=c(1/2)*100)
abline(h=c(1/3,2/3)*100)
```

Halton sequence

We wish to draw a sample of size $n=12$. To do this, we construct a random start Halton sequence. To do this, we select two random integers between 0 and U . In `SDraw`, U is defined in the function `maxU()` and defaults to 100,000,000. First random number is for the horizontal sequence, second is for the vertical.

```
library(SDraw)

## SDraw - Sample Draws (vers 2.1.8)
random.start <- floor(runif(2,max=maxU()+1))
random.start

## [1] 81436654 5817151
```

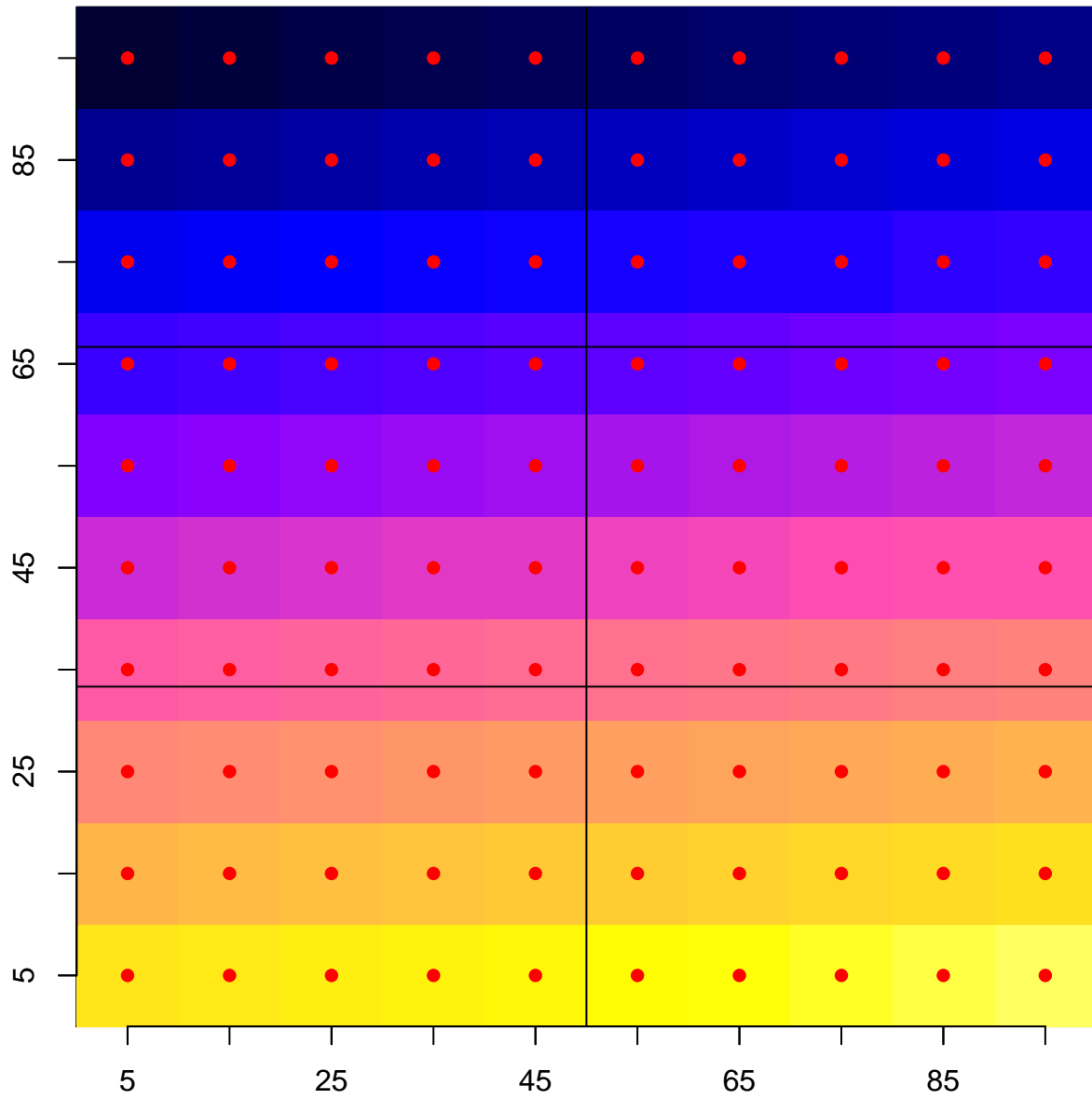


Figure 1: Example population with base $c(2,3)$ Halton boxes plotted. Population consists of 100 points in regular grid. Study area extent is 100 X 100 units. Point coordinates are 5, 15, 25, ..., 95

The random start Halton sequence starts at the random place, and proceeds for $n = 12$ indices. The Halton sequence produces numbers between 0 and 1, so these numbers need to be scaled to fit the study area.

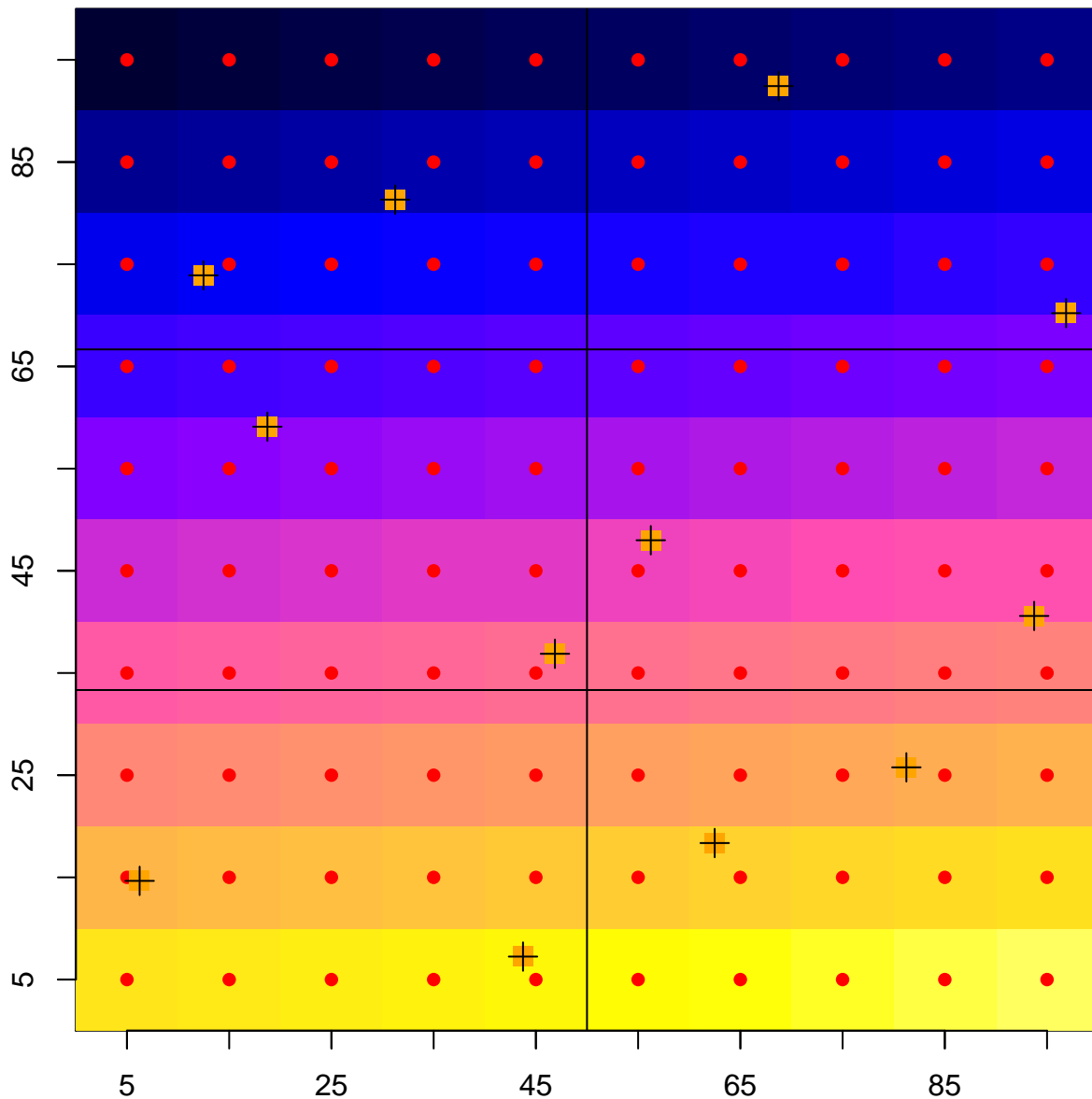
```
n <- 12
halt.seq <- halton(n, dim=2, start=random.start, bases=c(2,3))
# The unscaled random start Halton sequence
halt.seq
```

```
##           [,1]      [,2]
## [1,] 0.46864862 0.36877729
## [2,] 0.96864862 0.70211062
## [3,] 0.06239862 0.14655507
## [4,] 0.56239862 0.47988840
## [5,] 0.31239862 0.81322173
## [6,] 0.81239862 0.25766618
## [7,] 0.18739862 0.59099951
## [8,] 0.68739862 0.92433284
## [9,] 0.43739862 0.07248099
## [10,] 0.93739862 0.40581432
## [11,] 0.12489862 0.73914766
## [12,] 0.62489862 0.18359210
```

```
# The scaled random start Halton sequence
halt.seq <- halt.seq * 100
halt.seq
```

```
##           [,1]      [,2]
## [1,] 46.864862 36.877729
## [2,] 96.864862 70.211062
## [3,]  6.239862 14.655507
## [4,] 56.239862 47.988840
## [5,] 31.239862 81.322173
## [6,] 81.239862 25.766618
## [7,] 18.739862 59.099951
## [8,] 68.739862 92.433284
## [9,] 43.739862  7.248099
## [10,] 93.739862 40.581432
## [11,] 12.489862 73.914766
## [12,] 62.489862 18.359210
```

Next is a plot the scaled random start Halton points on the study area. Note in the plot that exactly 2 Halton points fall in each Halton box. This occurs because sample size (here, 12) is a multiple of a Halton cycle (here, 6).



The cyclic property of the Halton sequence extends to larger cycles, and it is this property that ensures spatial balance of BAS samples. The first cycle equating to the largest Halton boxes is the product of bases, i.e., $2(3) = 6$ here. Larger cycles are any higher power of 6, like 36, 216, etc. The second cycle equating to the next size smaller Halton boxes is $36 = 6*6$. Any sequence of 36 Halton points will have exactly one point in each of the 36 Halton boxes. Any sequence of $2(36) = 72$ Halton points will have exactly 2 points in each of the 36 Halton boxes. This is illustrated in the next plot.

BAS sampling

Units (pixels) are in the sample if a random start Halton sequence points falls inside the pixel. The following code determines pixel membership for all Halton points. In this case, determining sampled points is easy because the grid is square and regular. In general, one needs to apply point-in-polygon methods to determine sample membership. We demonstrate both methods below.

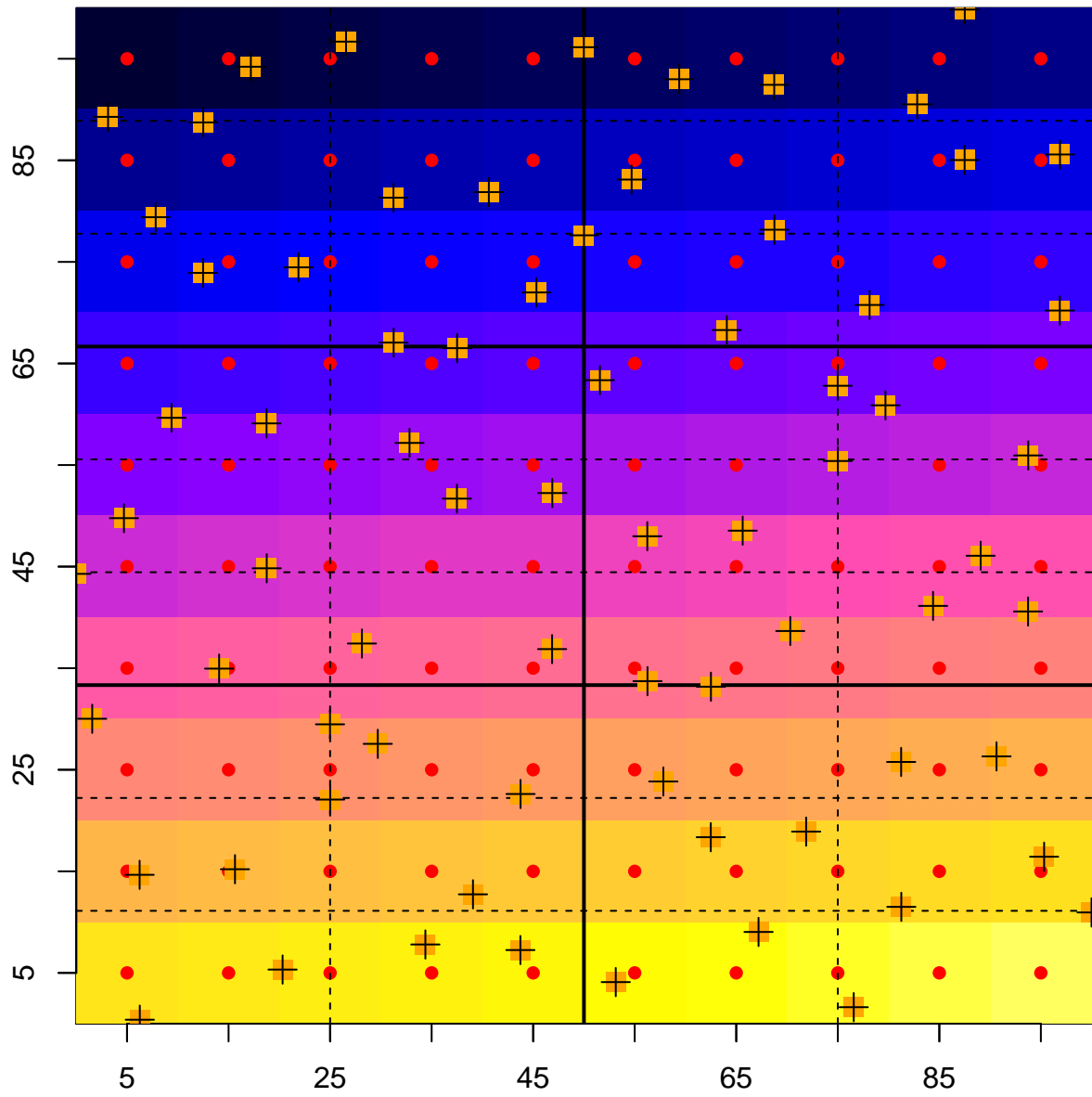


Figure 2: Study area with 72 Halton points. Exactly 2 Halton points appear in each of the 36 Halton boxes.

```

# Simple method to determine sampled points.
x.cell <- floor(halt.seq[,1]/10)
y.cell <- floor(halt.seq[,2]/10)

x.pt <- x.cell*10 + 5
y.pt <- y.cell*10 + 5

bas.samp <- data.frame(x=x.pt, y=y.pt)
bas.samp

##      x  y
## 1  45 35
## 2  95 75
## 3   5 15
## 4  55 45
## 5  35 85
## 6  85 25
## 7  15 55
## 8  65 95
## 9  45  5
## 10 95 45
## 11 15 75
## 12 65 15

```

The final BAS sample as a 'SpatialPoints' object:

```

bas.samp <- SpatialPoints(bas.samp)
bas.samp

## SpatialPoints:
##      x  y
## [1,] 45 35
## [2,] 95 75
## [3,]  5 15
## [4,] 55 45
## [5,] 35 85
## [6,] 85 25
## [7,] 15 55
## [8,] 65 95
## [9,] 45  5
## [10,] 95 45
## [11,] 15 75
## [12,] 65 15
## Coordinate Reference System (CRS) arguments: NA

```

The less illuminating but more general method of determining sample membership is to apply a point-in-polygon method.

```

halt.pts <- SpatialPoints(halt.seq)
keep <- over( halt.pts, popln.pixels)
bas.samp <- popln.pixels[keep$val,]
bas.samp

## Object of class SpatialPixelsDataFrame
## Object of class SpatialPixels
## Grid topology:

```

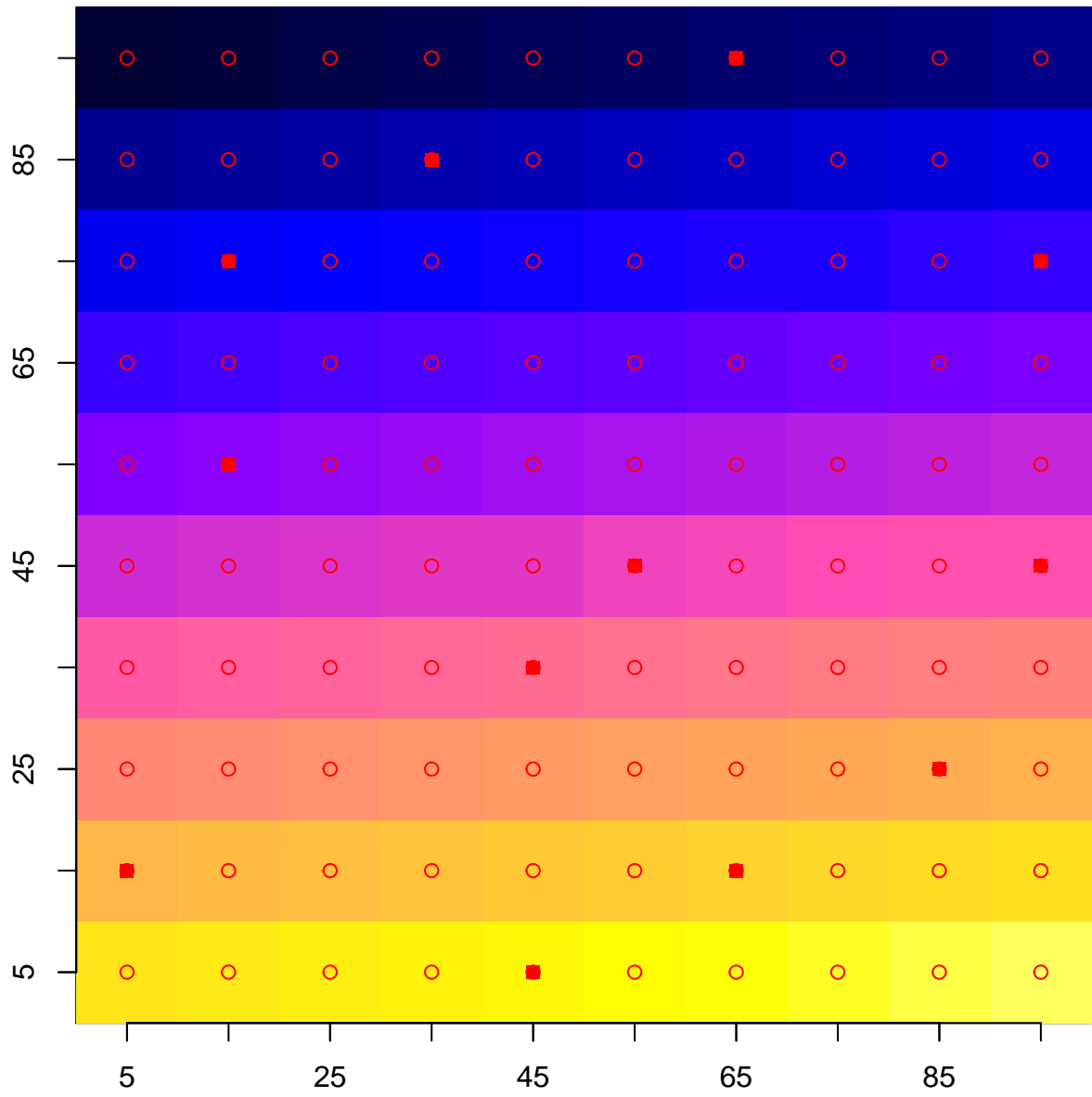


Figure 3: Example study area with $n = 12$ selected points (filled circles).

```
##   cellcentre.offset cellsize cells.dim
## s1                5      10      10
## s2                5      10      10
## SpatialPoints:
##      s1 s2
## [1,] 45 35
## [2,] 95 75
## [3,]  5 15
## [4,] 55 45
## [5,] 35 85
## [6,] 85 25
## [7,] 15 55
## [8,] 65 95
## [9,] 45  5
## [10,] 95 45
## [11,] 15 75
## [12,] 65 15
## Coordinate Reference System (CRS) arguments: NA
##
## Data summary:
##      val
## Min.   : 7.00
## 1st Qu.:28.00
## Median :58.00
## Mean   :53.17
## 3rd Qu.:79.50
## Max.   :95.00
```