Second Session

June 7, 2015

Table of Contents

1. Getting Familiar
   - Working with R
2. Functions
   - Functions
   - Arguments
3. Graphics
   - Histogram

Table of Contents

1. Getting Familiar
   - Working with R
2. Functions
   - Functions
   - Arguments
3. Graphics
   - Histogram

Last time...

- there is something called *working directory* telling R where on the computer it is supposed to search for file resp. save files (via the menu, `setwd()`, `getwd()`)
- there are numbers, characters and logical values (TRUE, FALSE)
- there is the `c()` function to create vectors
- there are data frames which are the R object most similar to data tables known from SPSS and Excel
- missing values are coded as `NA`
- with the `$` I can access columns of such a *data frame*
- there are indices (numbers, characters, logical values)
- everything can be assigned to a variable

Learned last time... ✓

- ... and using the tab key!!!!

Table of Contents

1 Getting Familiar
- Working with R

2 Functions
- Functions
- Arguments

3 Graphics
- Histogram

Functions

- we've seen them already, we've used them already
- they are everywhere, they do all the work

FUNCTIONS

Functions and Arguments

- functions are just like what you remember from school
- most functions are in the following form:
  `f(argument1,argument2,...)`

Functions and Arguments

- the arguments are named, try

```
> log(x=64,base=4)
[1] 3
> log(base=4,x=64)
[1] 3
```

Functions - Arguments

- arguments also have a predefined order (which you can explore using the `?command`)
- if you do not use names, they are used in this predefined order, try

```
> log(64,4)
[1] 3
> log(4,64)
[1] 0.3333333
> ?log
```

Functions - Arguments

- arguments can be required
- or they can be optional
- they can have default values (again check `?command`)

```
> log(64,4)
[1] 3
> log(64)
[1] 4.158883
> ?log
```

Table of Contents

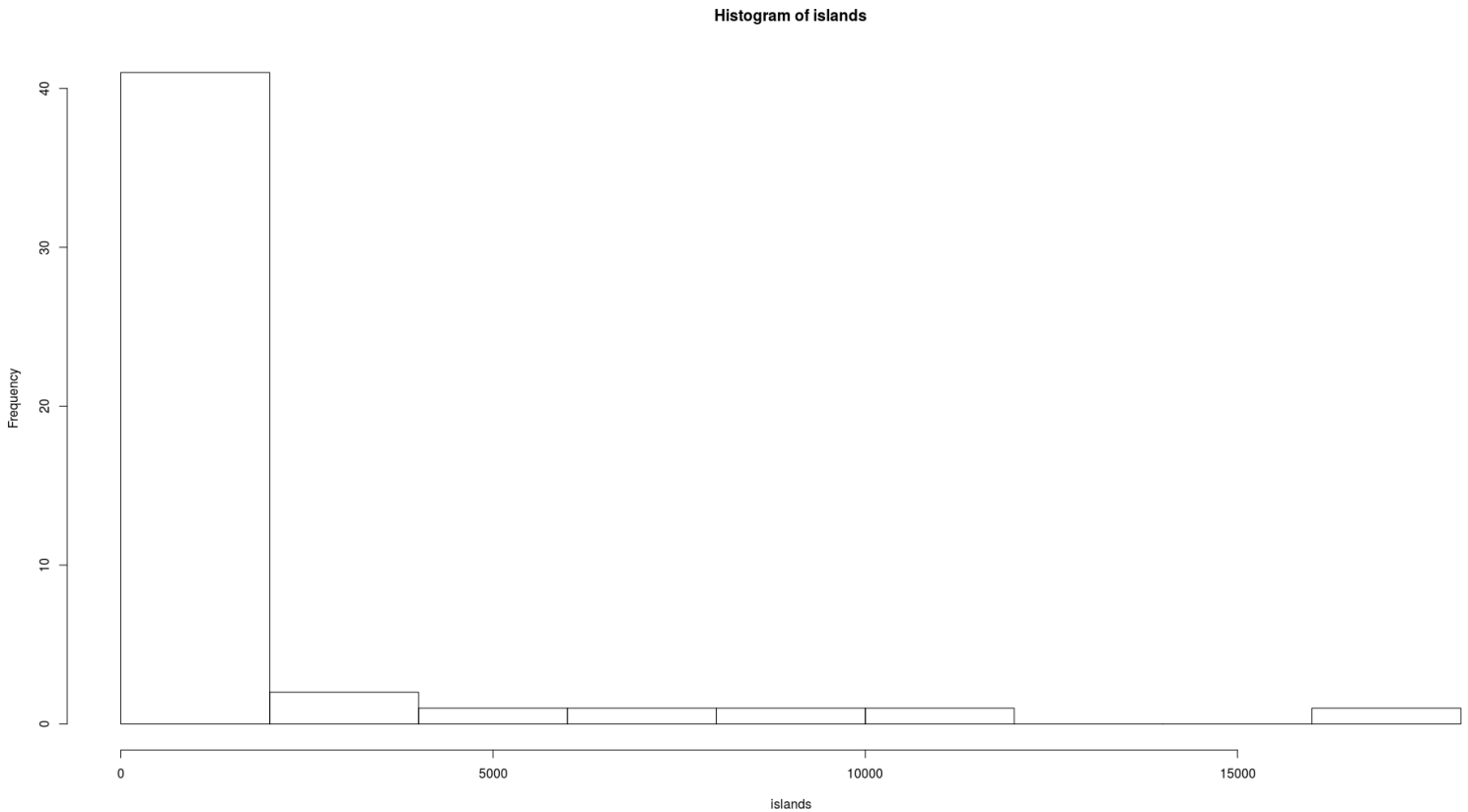
1. Getting Familiar
   - Working with R
2. Functions
   - Functions
   - Arguments
3. Graphics
   - Histogram

Histogram

- a histogram is a set of contiguously drawn bars showing a frequency distribution
- bars are drawn for each group (interval) such that the area is proportional to the frequency in that group
- variable values are plotted on the horizontal (x-axis)
- frequencies on the vertical axis (y-axis)
- the r command (basic graphics) is `hist()`

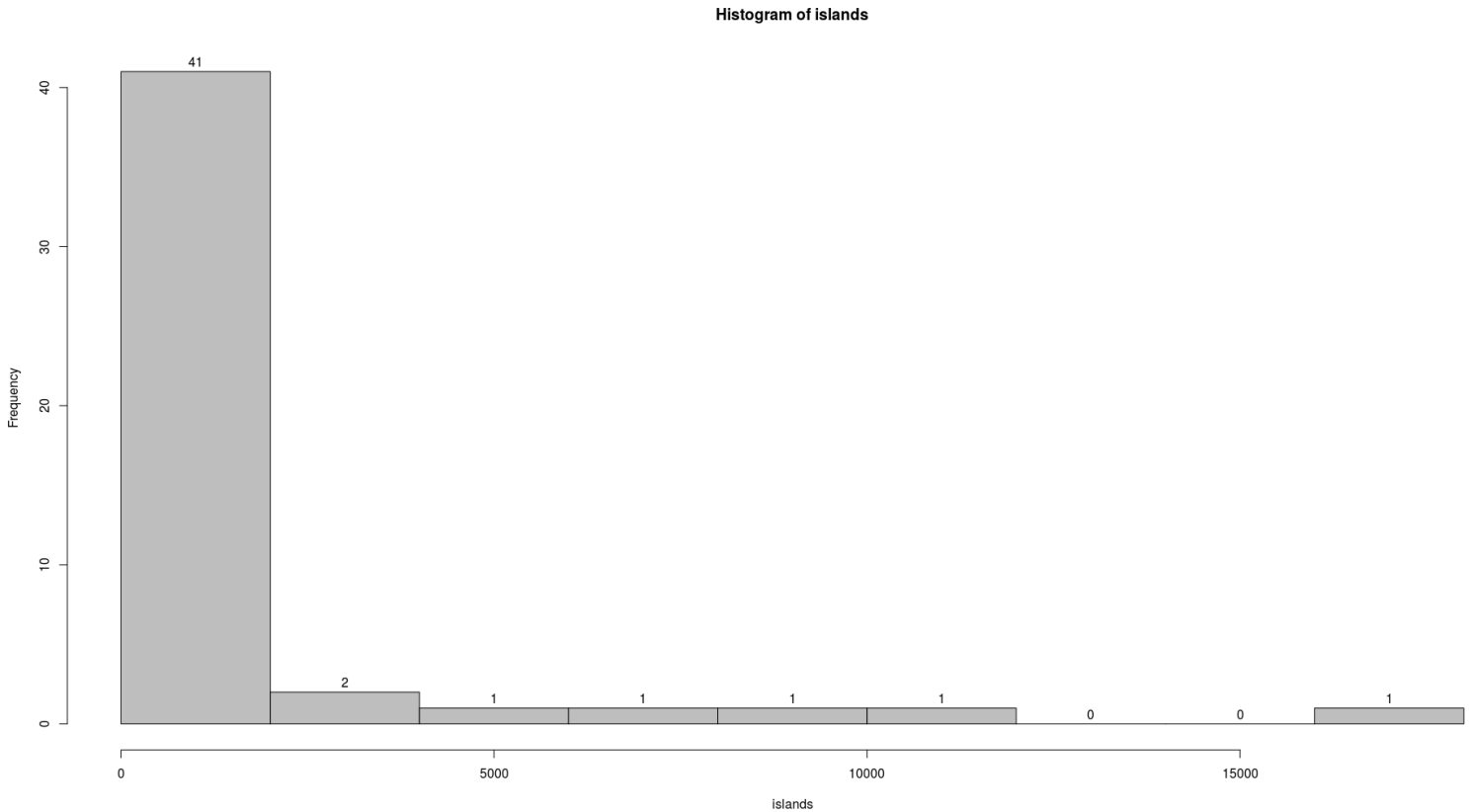
Histogram

```
> hist(islands)
```



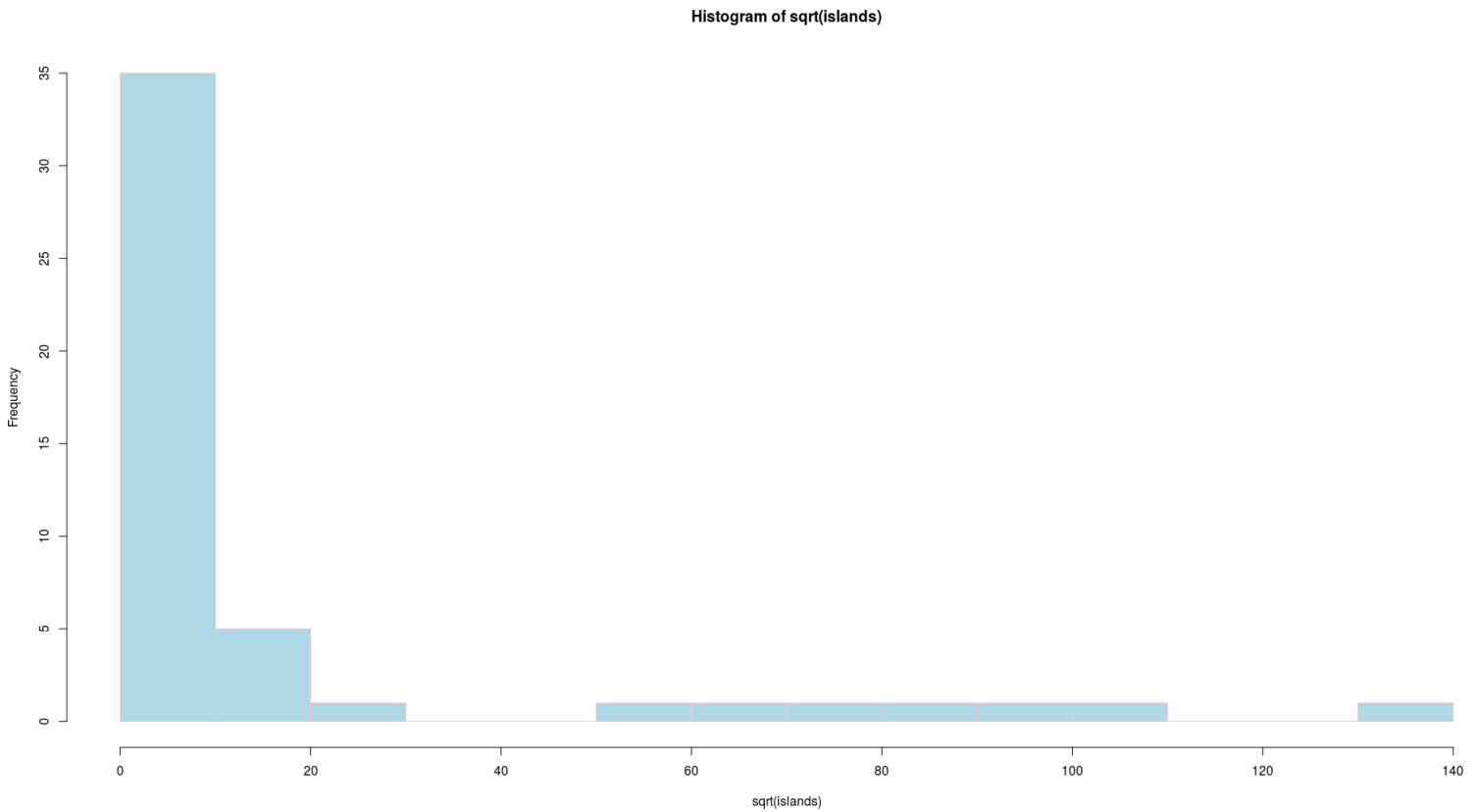
Histogram

```
> hist(islands, col = "gray", labels = TRUE)
```



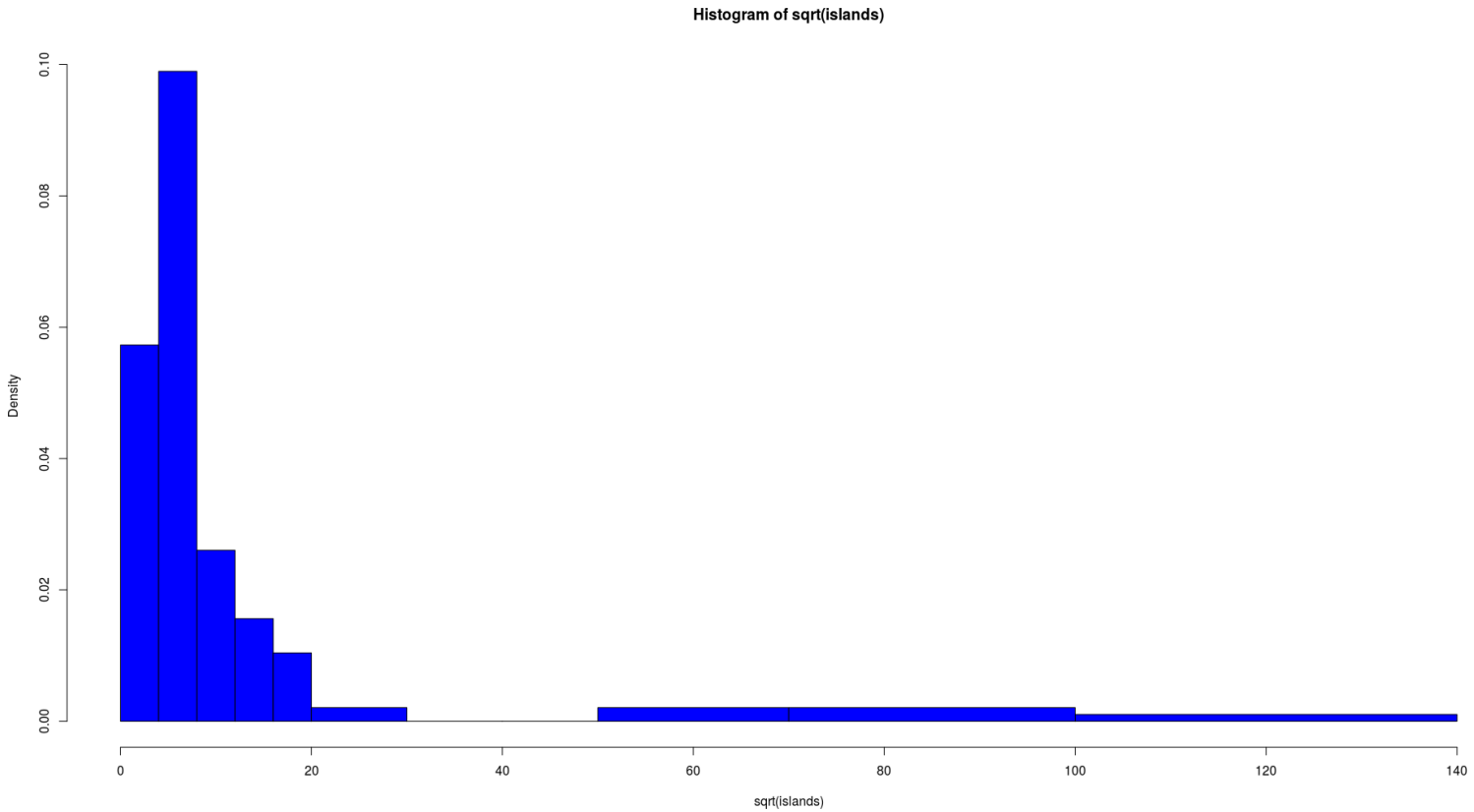
Histogram

```
> hist(sqrt(islands), breaks = 12, col = "lightblue",
+   border = "pink")
```



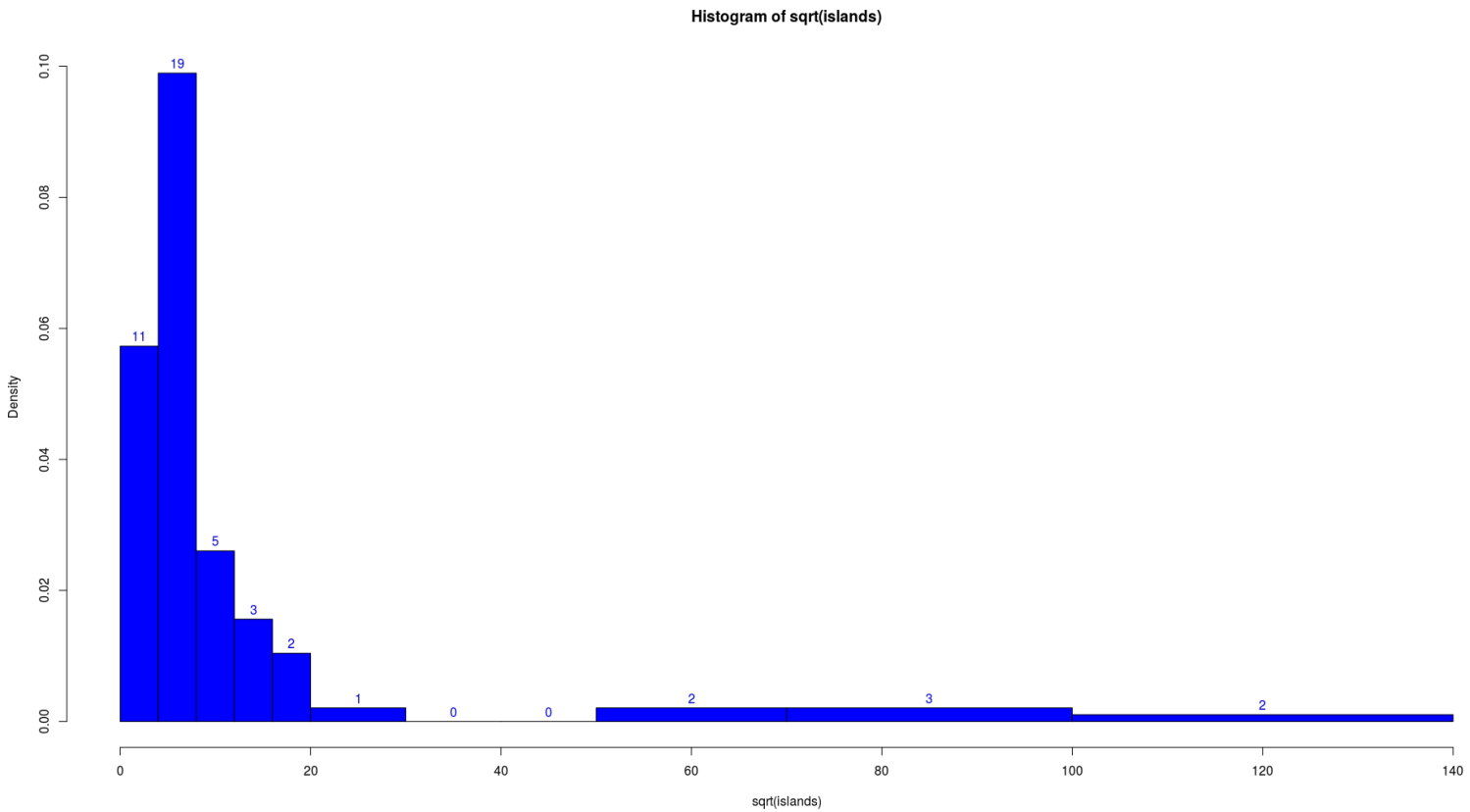
Histogram

```
r <- hist(sqrt(islands),
+      breaks = c(4*0:5, 10*3:5, 70, 100, 140),
+      col = "blue1")
```



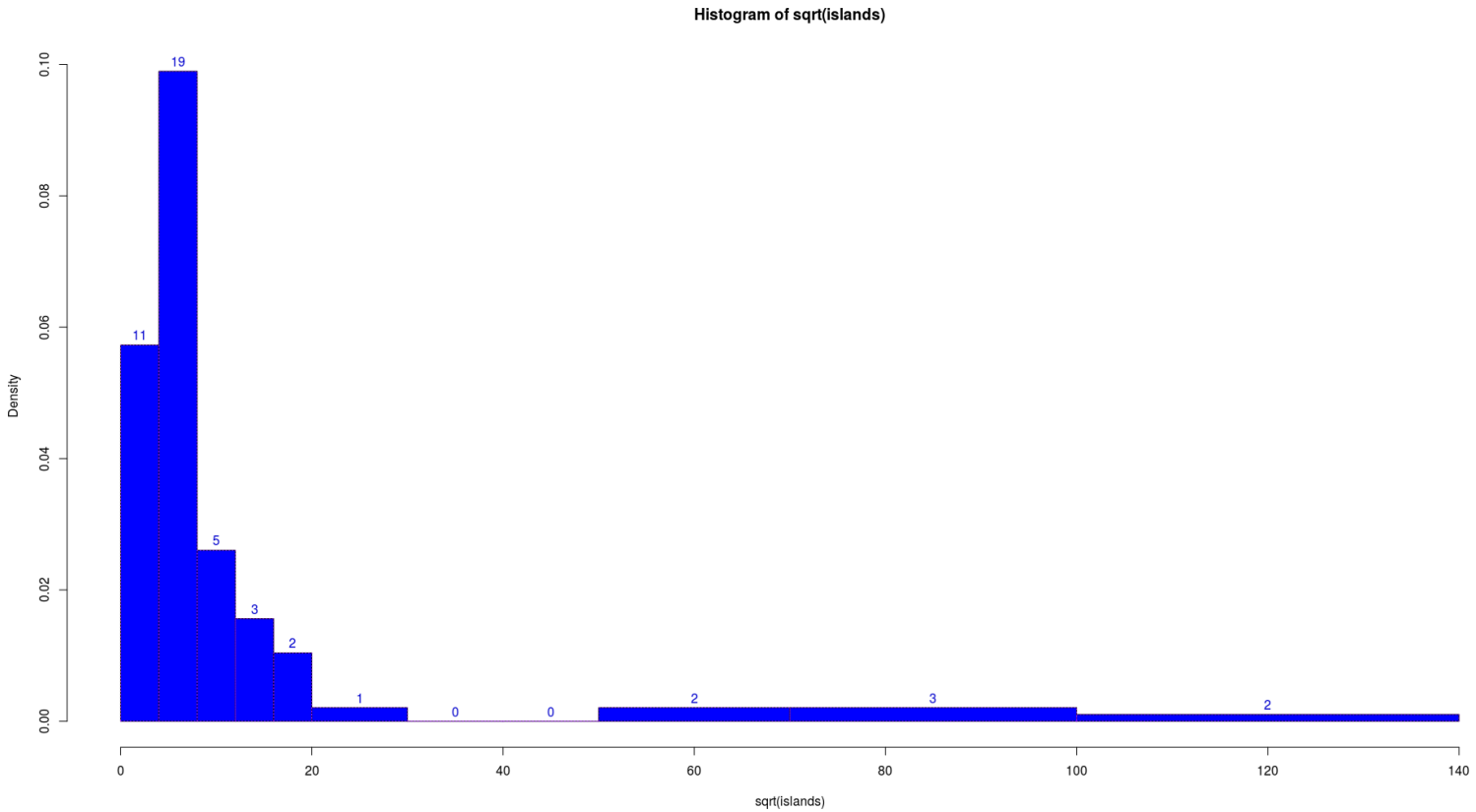
Histogram

```
> text(r$mids, r$density, r$counts,
+   adj = c(.5, -.5), col = "blue3")
```



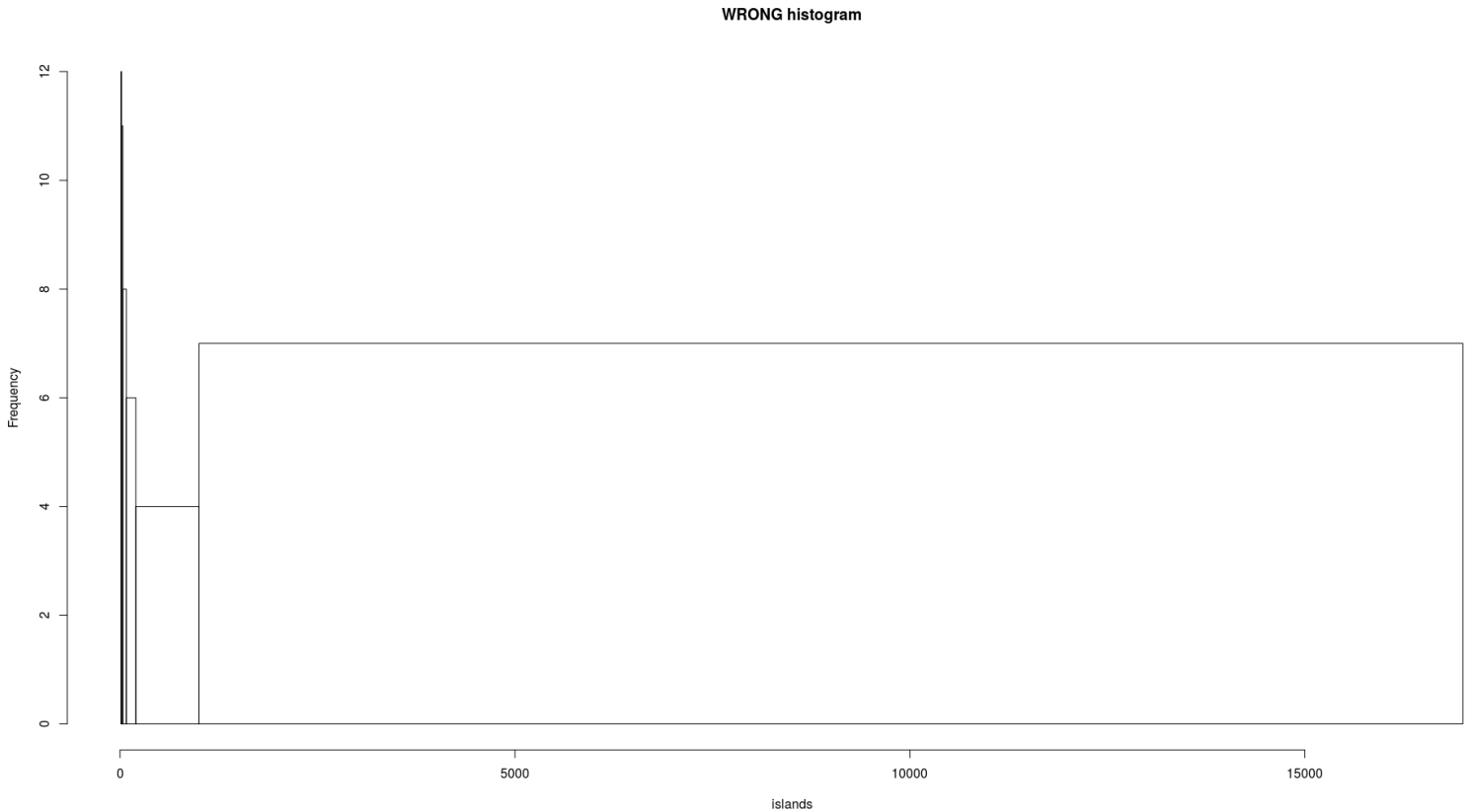
Histogram

```
> lines(r, lty = 3, border = "purple") # -> lines
```



Histogram

```
> hist(islands, breaks = c(12,20,36,80,200,1000,17000),
+   freq = TRUE, main = "WRONG histogram")
```



Histogram

```
> par(mfrow=c(1,2))
> set.seed(14)
> x <- rchisq(100, df = 4)
> qqplot(x, qchisq(ppoints(x), df = 4))
> abline(0, 1, col = 2, lty = 2)
> hist(x, freq = FALSE, ylim = c(0, 0.2))
> curve(dchisq(x, df = 4), col = 2,
+  lty = 2, lwd = 2, add = TRUE)
```