HOWTO

Drawing a **histogram** from a frequency table with R

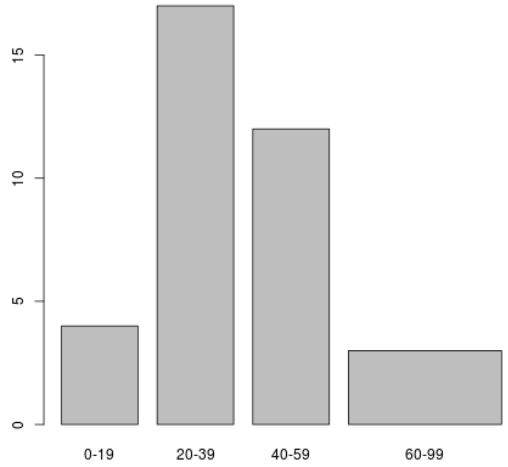
Frequency table data can be displayed in a boxplot, which is equivalent to the histogram for that data.

Example:

Using the same frequency table as above, we call the barplot function in R:

```
> barplot(height=c(4,17,12,3),width=c(1,1,1,2),names=c("0-19","20-39","40-59","60-99"))
```

The height vector contains the calculated frequency densities, the width vector contains the widths of the ranges in 'units of range' and the names vector contains the range labels.



HOWTO

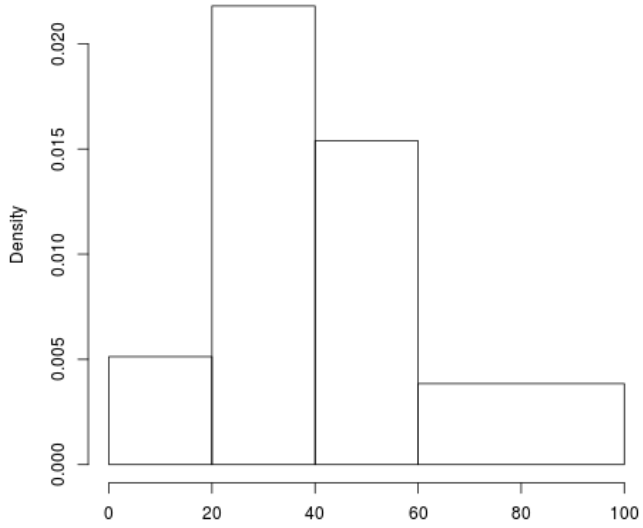
Drawing a **histogram** from a dataset with R

A function is available in R for drawing a histogram directly from a vector containing some attribute's values.

Example:

Let's 'make up' some data that fits the previous example. We feed the data into the R hist() function, with the resulting diagram equivalent to the one obtained from the frequency table.

```
> hist(c(3,5,7,17,21,21,23,25,25,28,30,31,32,33,33,33,34,36,37,38,39,41,44,44,45,47,49,
50,50,52,55,56,57,62,66,70,76,84,91),breaks=c(0,20,40,60,100))
```



NOTE: The values on the y-axis are chosen so that the area under the entire histogram is equal to 1 (the 'unit of range' is 1/39, where 39 is the number of values in the set). However, the proportions remain the same as in the histogram created from the frequency table in the previous HOWTO panel.

HOWTO

Finding the **range** of a dataset with R

Use the following R command:

```
> max(DATASET) - min(DATASET)
```

Example:

```
> dataset <- c(2, 3, 4, 4, 5, 5, 5, 7, 9, 10, 10, 11, 12, 12, 15)
> max(dataset) - min(dataset)
[1] 13
```

The value given for the range by R is 13.

HOWTO

Finding the **inter-quartile range (IQR)** with R

```
> IQR(DATASET)
```

Example:

```
> IQR(c(2, 3, 4, 4, 5, 5, 5, 7, 9, 10, 10, 11, 12, 12, 15))

 [1] 6
```

NOTE: The value returned by R (6) differs from that calculated manually (7). This is because there are several methods for determining the IQR. All these are valid and the difference in results is not significant in the case of large datasets, which are the type that are encountered in real-life statistics.