R boxplot() to Create Box Plot (With Numerous Examples)

R boxplot() to Create Box Plot (With Numerous Examples)

In this article, you will learn to create whisker and box plot in R programming. You will also learn to draw multiple box plots in a single plot.

In R, boxplot (and whisker plot) is created using the `boxplot()` function.

The `boxplot()` function takes in any number of numeric vectors, drawing a boxplot for each vector.

You can also pass in a list(or data frame) with numeric vectors as its components. Let us use the built-in dataset `airquality` which has "Daily air quality measurements in New York, May to September 1973."-R documentation.

```
> str(airquality)
'data.frame':   153 obs. of  6 variables:
 $ Ozone  : int  41 36 12 18 NA 28 23 19 8 NA ...
 $ Solar.R: int  190 118 149 313 NA NA 299 99 19 194 ...
 $ Wind   : num  7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 ...
 $ Temp   : int  67 72 74 62 56 66 65 59 61 69 ...
 $ Month  : int  5 5 5 5 5 5 5 5 5 5 ...
 $ Day    : int  1 2 3 4 5 6 7 8 9 10 ...
```

Let us make a boxplot for the ozone readings.

```
boxplot(airquality$Ozone)
```
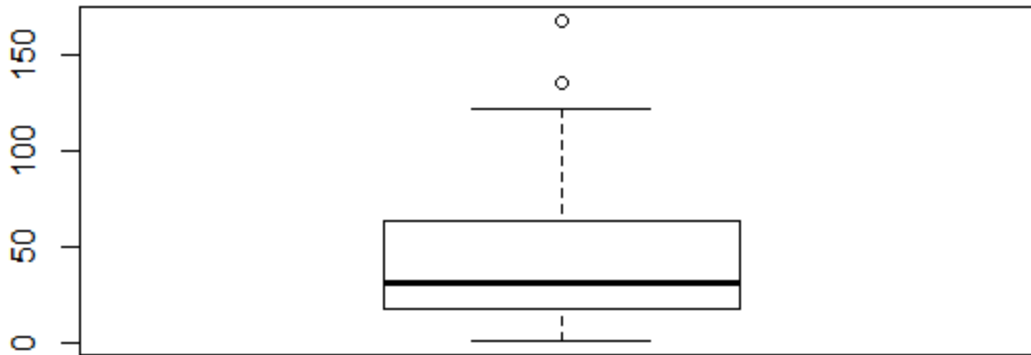
We can see that data above the median is more dispersed. We can also notice two outliers at the higher extreme.

We can pass in additional parameters to control the way our plot looks. You can read about them in the help section `?boxplot` .

Some of the frequently used ones are, `main` -to give the title, `xlab` and `ylab` -to provide labels for the axes, `col` to define color etc.

Additionally, with the argument `horizontal = TRUE` we can plot it horizontally and with `notch = TRUE` we can add a notch to the box.

R boxplot() to Create Box Plot (With Numerous Examples)

```
boxplot(airquality$Ozone,
main = "Mean ozone in parts per billion at Roosevelt Island",
xlab = "Parts Per Billion",
ylab = "Ozone",
col = "orange",
border = "brown",
horizontal = TRUE,
notch = TRUE
)
```


Mean ozone in parts per billion at Roosevelt Island
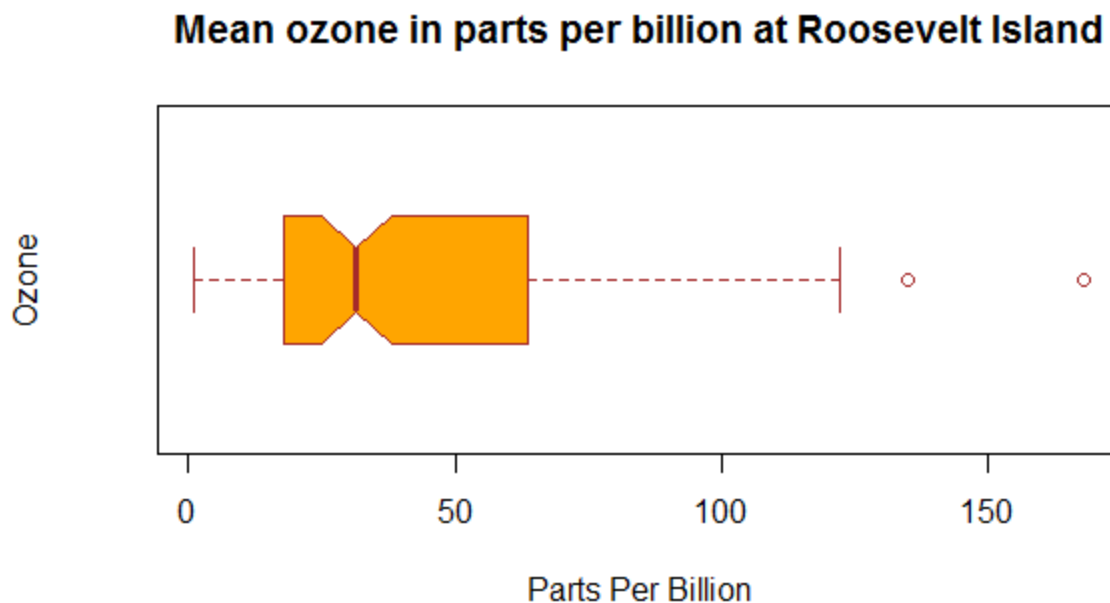
Return Value of boxplot()

The `boxplot()` function returns a list with 6 components shown as follows.

R boxplot() to Create Box Plot (With Numerous Examples)

```
> b <- boxplot(airquality$Ozone)
> b
$stats
[,1]
[1,]   1.0
[2,]  18.0
[3,]  31.5
[4,]  63.5
[5,] 122.0
attr(,"class")
1
"integer"
$n
[1] 116
$conf
[,1]
[1,] 24.82518
[2,] 38.17482
$out
[1] 135 168
$group
[1] 1 1
$names
[1] "1"
```

As we can see above, a list is returned which has stats -having the position of the upper/lower extremes of the whiskers and box along with the median,

- n -the number of observation the boxplot is drawn with (notice that NA 's are not taken into account)
- conf -upper/lower extremes of the notch, out -value of the outliers
- group -a vector of the same length as out whose elements indicate to which group the outlier belongs and
- names -a vector of names for the groups.

R boxplot() to Create Box Plot (With Numerous Examples)

Multiple Boxplots

We can draw multiple boxplots in a single plot, by passing in a list, data frame or multiple vectors.

Let us consider the Ozone and Temp field of airquality dataset. Let us also generate normal distribution with the same mean and standard deviation and plot them side by side for comparison.

```
# prepare the data
ozone <- airquality$Ozone
temp <- airquality$Temp
# gererate normal distribution with same mean and sd
ozone_norm <- rnorm(200,mean=mean(ozone, na.rm=TRUE), sd=sd(ozone, na.rm=TRUE))
temp_norm <- rnorm(200,mean=mean(temp, na.rm=TRUE), sd=sd(temp, na.rm=TRUE))
```

Now we us make 4 boxplots with this data. We use the arguments at and names to denote the place and label.

```
boxplot(ozone, ozone_norm, temp, temp_norm,
main = "Multiple boxplots for comparision",
at = c(1,2,4,5),
names = c("ozone", "normal", "temp", "normal"),
las = 2,
col = c("orange","red"),
border = "brown",
horizontal = TRUE,
notch = TRUE
)
```

Multiple boxplots for comparision
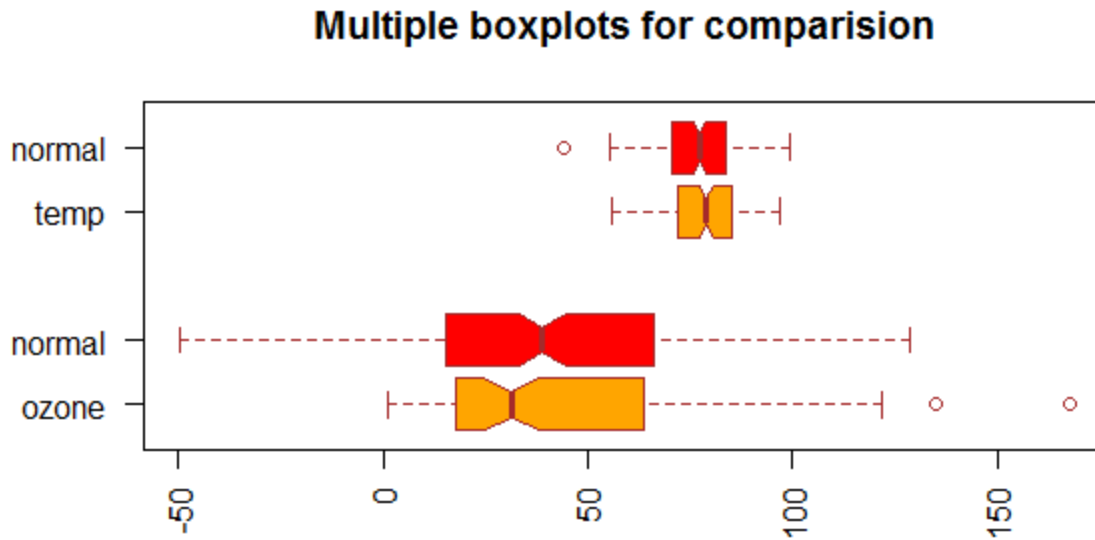
Boxplot form Formula

The function `boxplot()` can also take in formulas of the form `y~x` where, `y` is a numeric vector which is grouped according to the value of `x`.

For example, in our dataset `airquality`, the `Temp` can be our numeric vector. Month can be our grouping variable, so that we get the boxplot for each month separately. In our dataset, month is in the form of number (1=January, 2-Febuary and so on).

```
boxplot(Temp~Month,
data=airquality,
main="Different boxplots for each month",
xlab="Month Number",
ylab="Degree Fahrenheit",
col="orange",
border="brown"
)
```


Different boxplots for each month
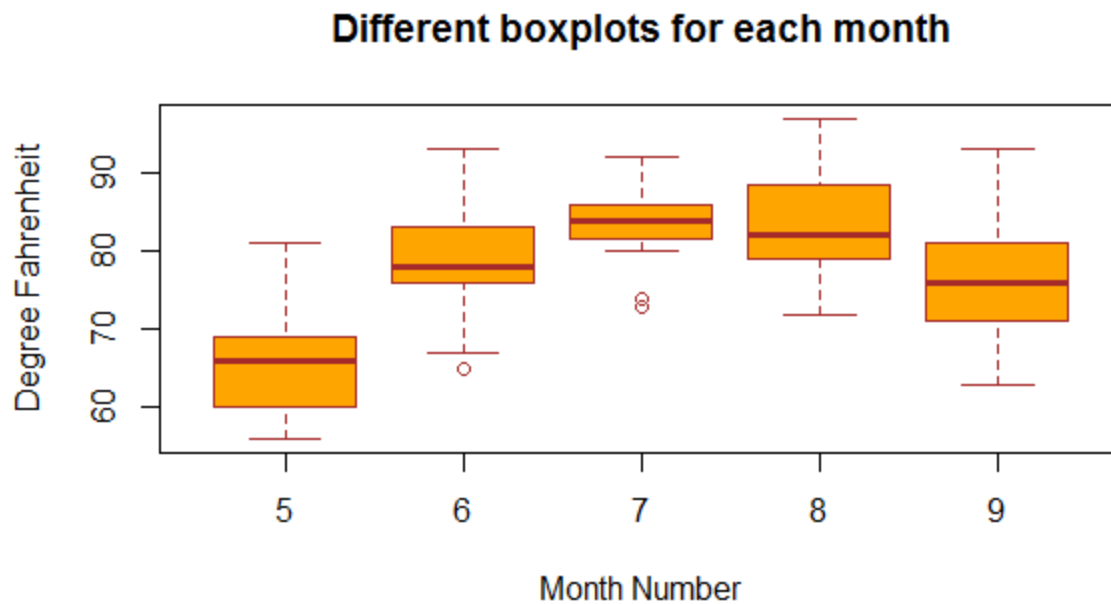
It is clear from the above figure that the month number 7 (July) is relatively hotter than the rest.