

How do I remove outliers in boxplots in R?

Authored by
stats writer

April 18, 2024

RECOMMENDED CITATION

stats writer (2024). *How do I remove outliers in boxplots in R?*. PSYCHOLOGICAL SCALES.
Retrieved from <https://scales.arabpsychology.com/?p=136733>

Boxplots are a commonly used graphical tool in statistics for visualizing the distribution of numerical data. Outliers, or extreme values, can greatly affect the interpretation of the data in a boxplot. In R, outliers can be identified and removed using various techniques such as the Tukey method, which defines outliers as values that fall outside the range of 1.5 times the interquartile range. To remove outliers in boxplots in R, one can use the "boxplot.stats" function to compute the statistics of the data and then use the "boxplot" function to create a new boxplot without the outliers. This process can help to better understand the central tendency and spread of the data and improve the accuracy of statistical analysis.

Remove Outliers in Boxplots in R

Occasionally you may want to remove outliers from boxplots in R.

This tutorial explains how to do so using both base R and ggplot2.

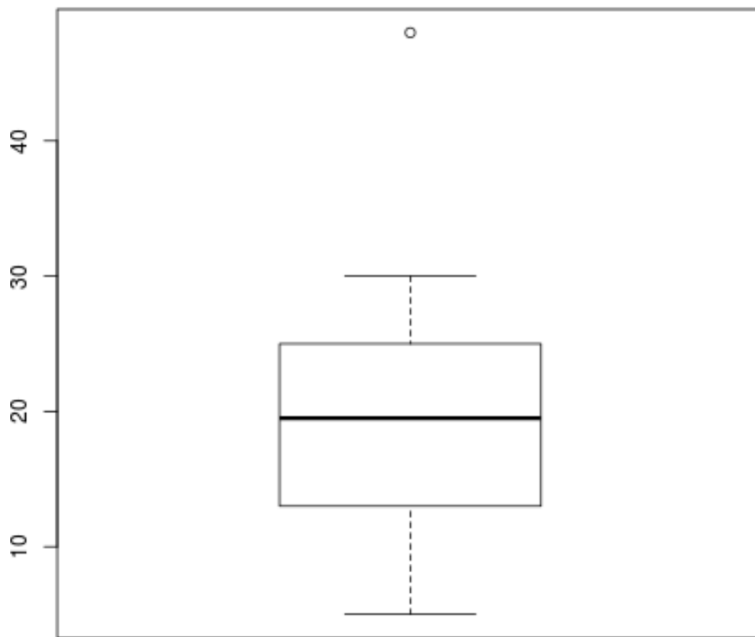
Remove Outliers in Boxplots in Base R

Suppose we have the following dataset:

```
data <- c(5, 8, 8, 12, 14, 15, 16, 19, 20, 22, 24, 25, 25, 26, 30, 48)
```

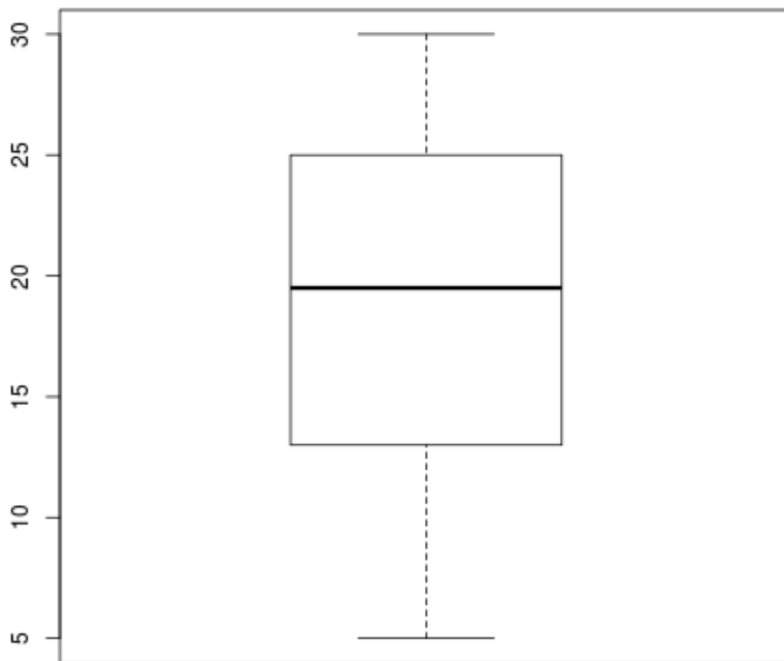
The following code shows how to create a boxplot for this dataset in base R:

```
boxplot(data)
```



To remove the outliers, you can use the argument `outline=FALSE`:

`boxplot(data, outline=FALSE)`



Remove Outliers in Boxplots in ggplot2

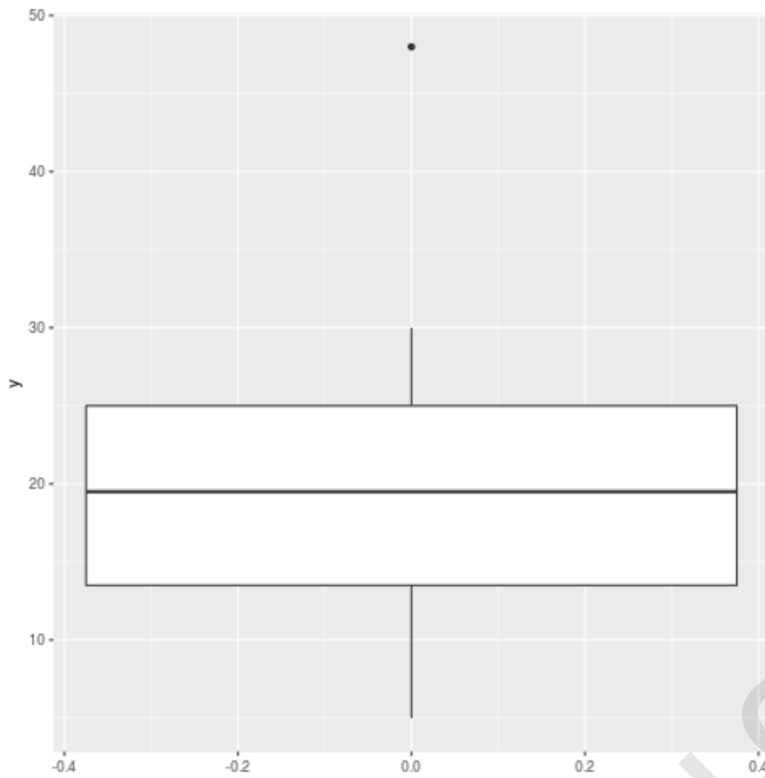
Suppose we have the following dataset:

```
data <- data.frame(y=c(5, 8, 8, 12, 14, 15, 16, 19, 20, 22, 24, 25, 25, 26, 30, 48))
```

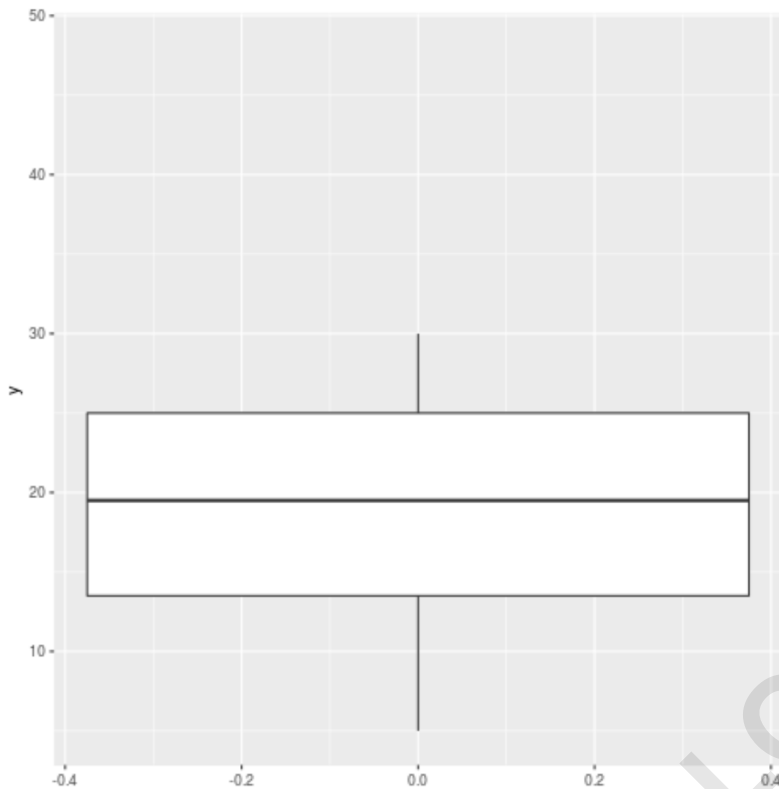
The following code shows how to create a boxplot using the ggplot2 visualization library:

```
library(ggplot2)
```

```
ggplot(data, aes(y=y)) +  
geom_boxplot()
```



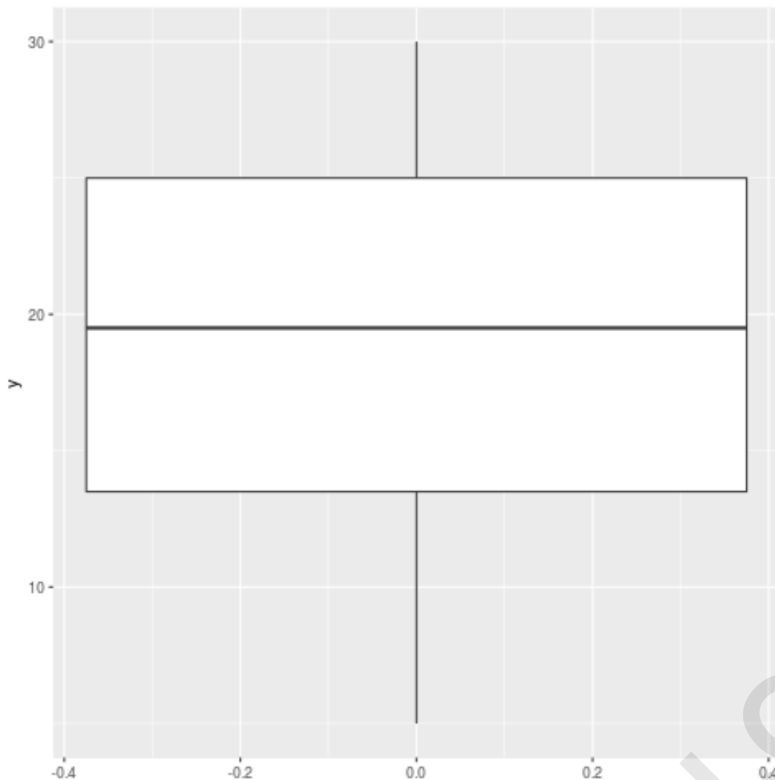
```
ggplot(data, aes(y=y)) +  
geom_boxplot(outlier.shape = NA)
```



Notice that ggplot2 does not automatically adjust the y-axis.

To adjust the y-axis, you can use `coord_cartesian`:

```
ggplot(data, aes(y=y)) +  
geom_boxplot(outlier.shape = NA) +  
coord_cartesian(ylim=c(5, 30))
```



The y-axis now ranges from 5 to 30, just as we specified using the `yylim()` argument.

The following tutorials explain how to perform other common operations in `ggplot2`:

[How to Set Axis Limits in ggplot2](#)

[How to Create Side-by-Side Plots in ggplot2](#)