GoodReads_100k_books

Deepika Jayanth

2022-06-29

```
data<-read.csv("C:/Users/DEEPIKA/OneDrive/Desktop/EDA_PROJECT/archive/GoodReads_100k_books1.csv")
```

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

``

Including Plots

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

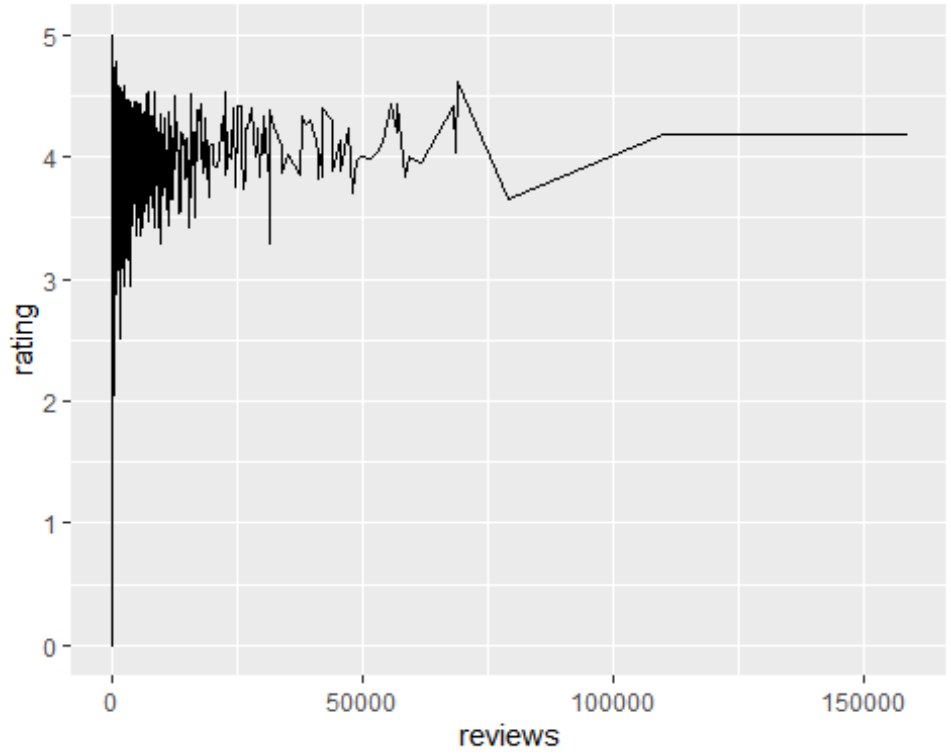
loading library to access various functions related to graph

```
library(ggplot2)

library(gcookbook)
```

making a basic line graph.

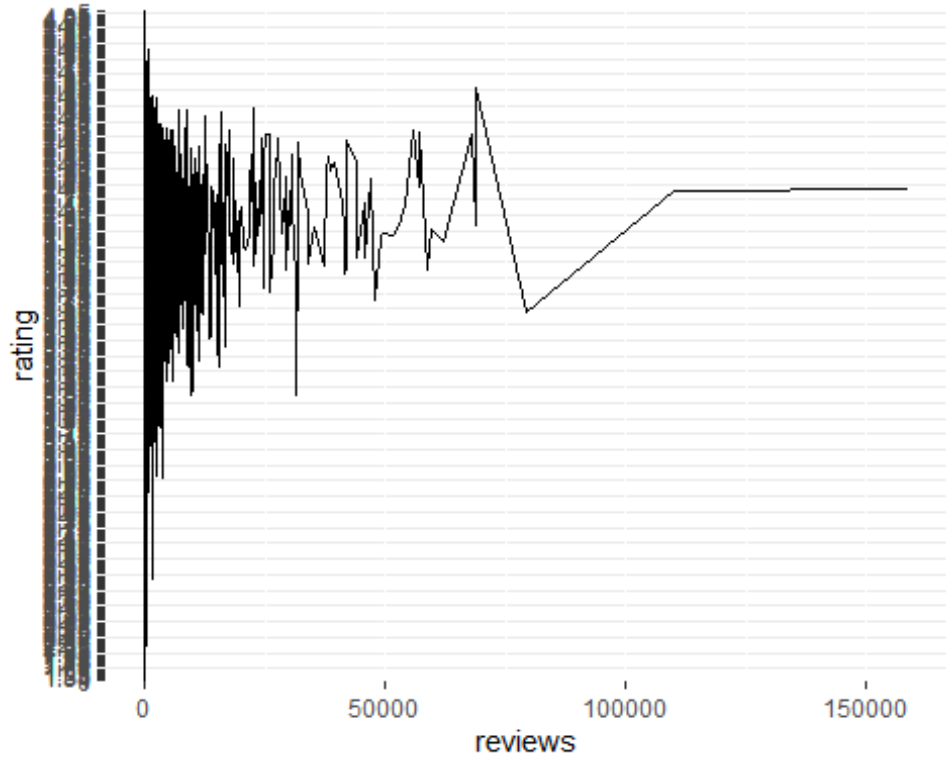
```
ggplot(data, aes(x =reviews, y = rating)) +
 geom_line()
```



Make a copy of the data

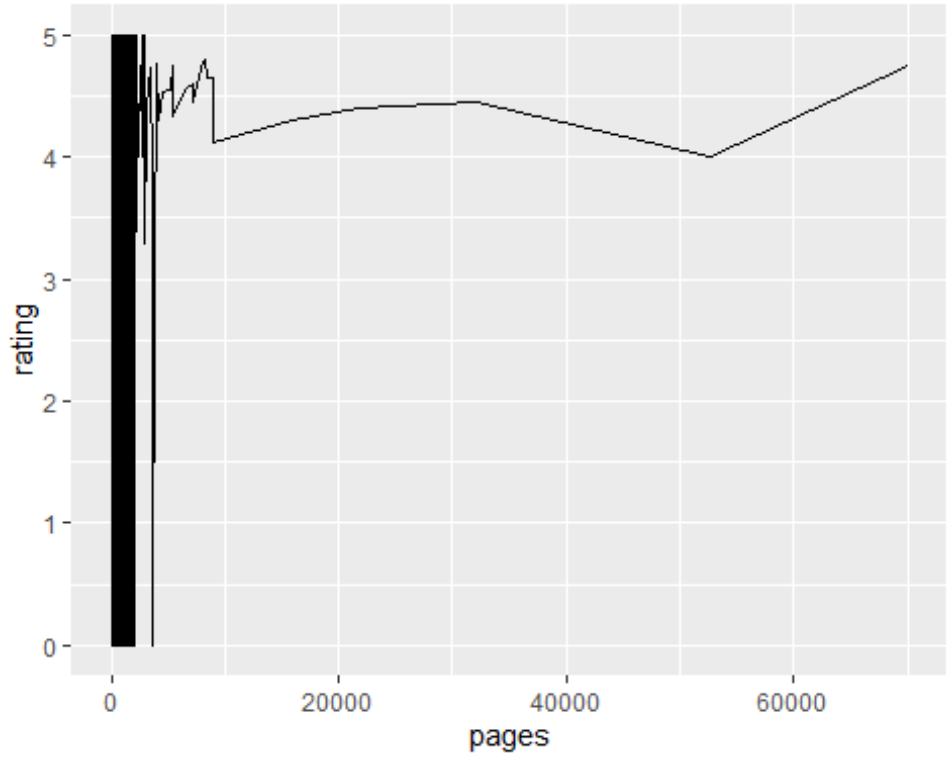
```
data1 <- data # Make a copy of the data
data1$rating <- factor(data1$rating)

ggplot(data1, aes(x = reviews, y = rating, group = 1)) +
 geom_line()
```

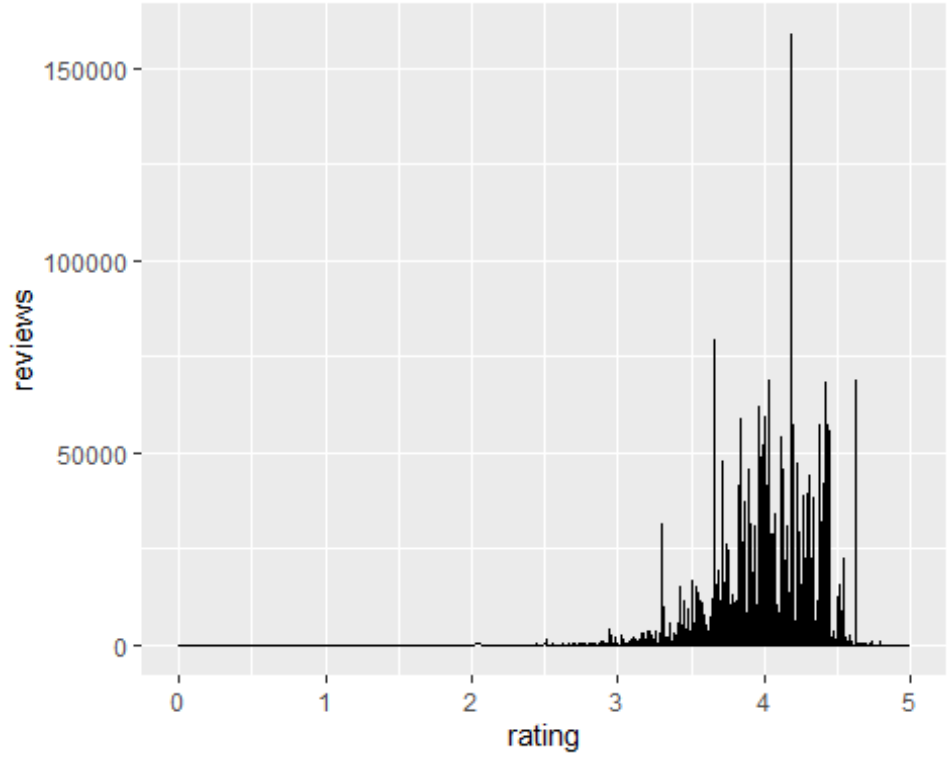


This will also give same result

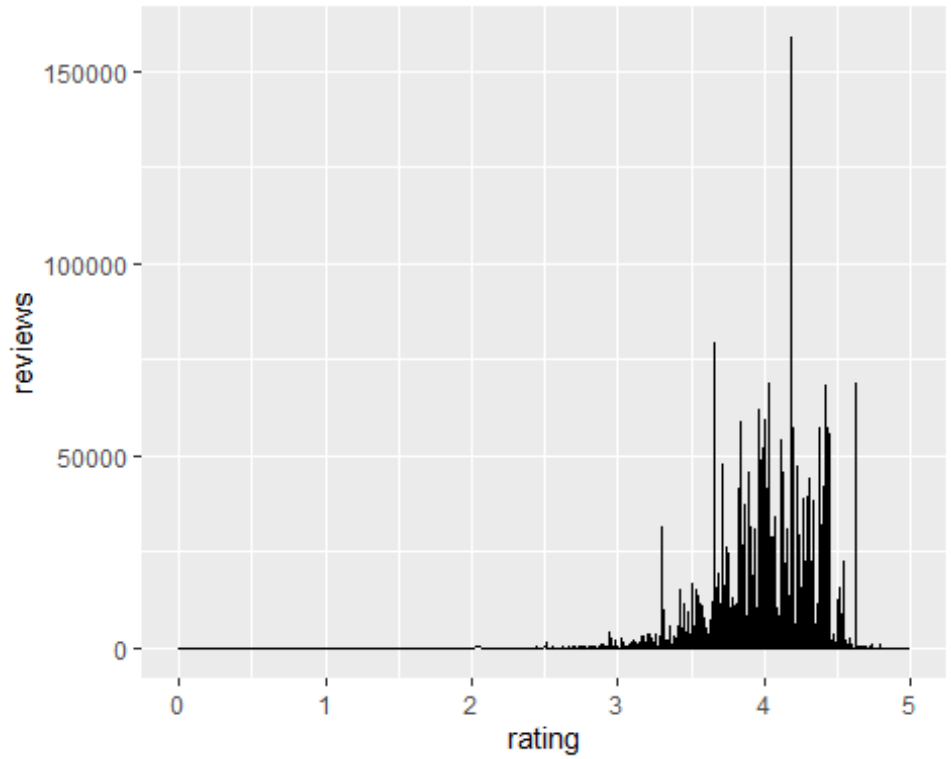
```
ggplot(data, aes(x = pages, y = rating)) +
 geom_line() +
 ylim(0, max(data$rating))
```




```
ggplot(data, aes(x = rating, y = reviews)) +
 geom_line() +
 expand_limits(y = 0)
```

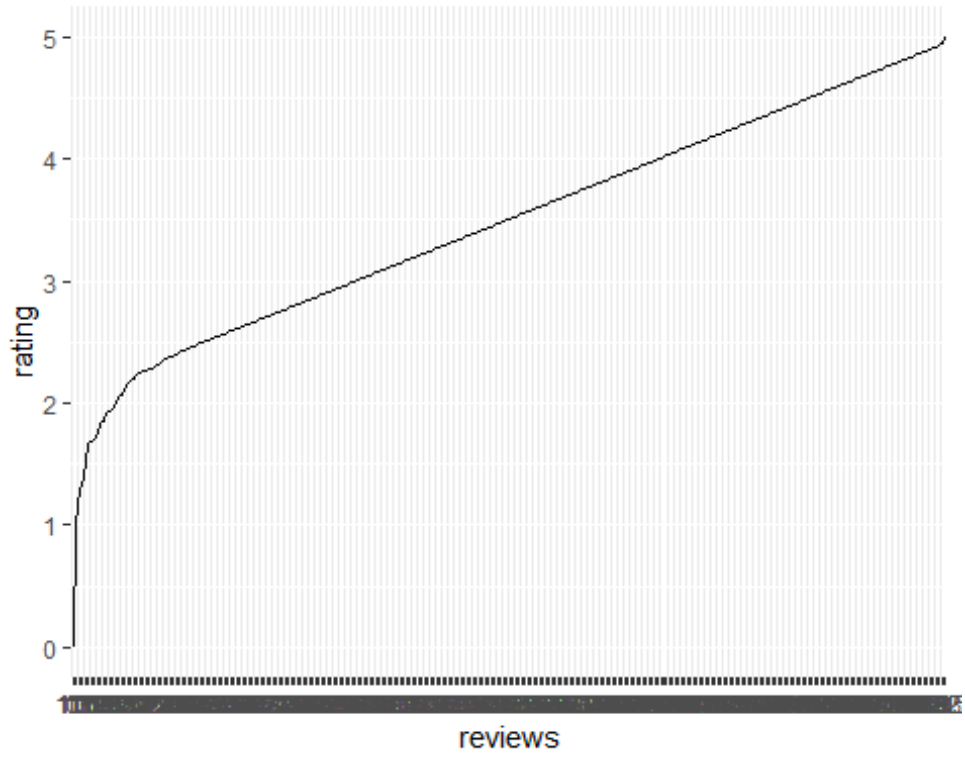


```
ggplot(data, aes(x =rating, y =reviews )) +
geom_line()
```

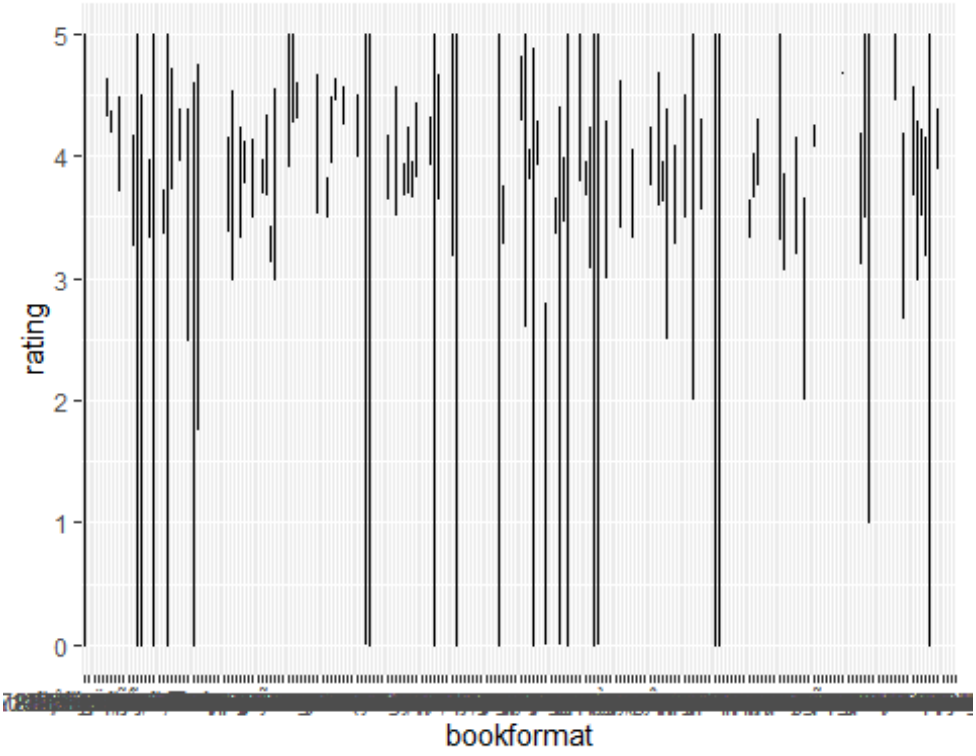


Make a copy of the data

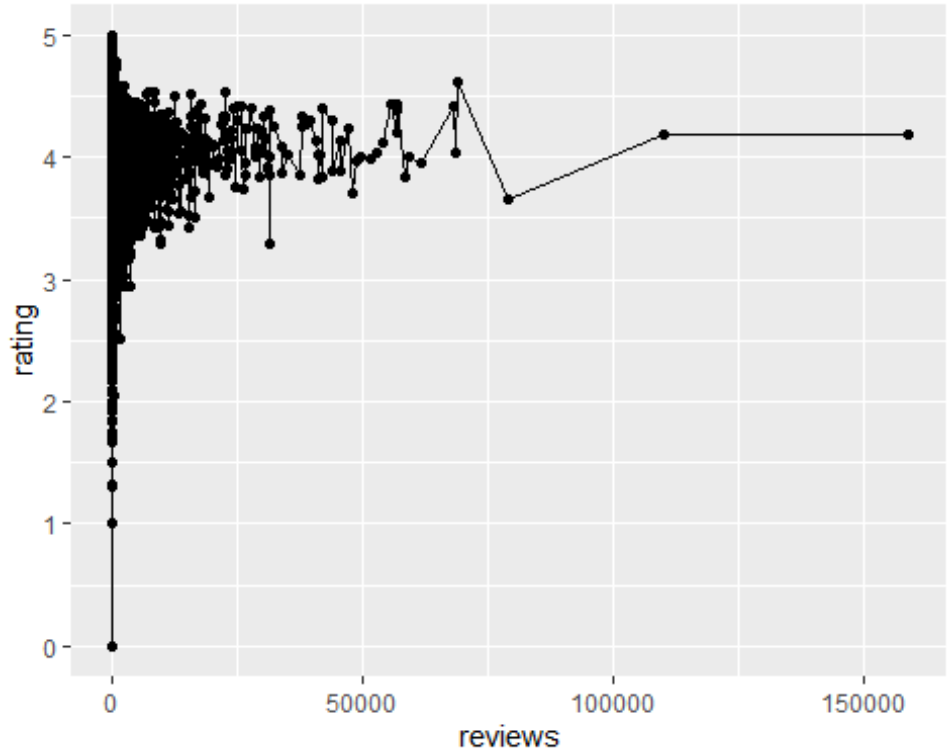
```
data1 <- data
data1$reviews <- factor(data1$rating)
ggplot(data1, aes(x = reviews, y = rating, group = 1)) +
geom_line()
```




```
ggplot(data, aes(x = bookformat, y = rating)) +
 geom_line() +
 expand_limits(y = 0)
```

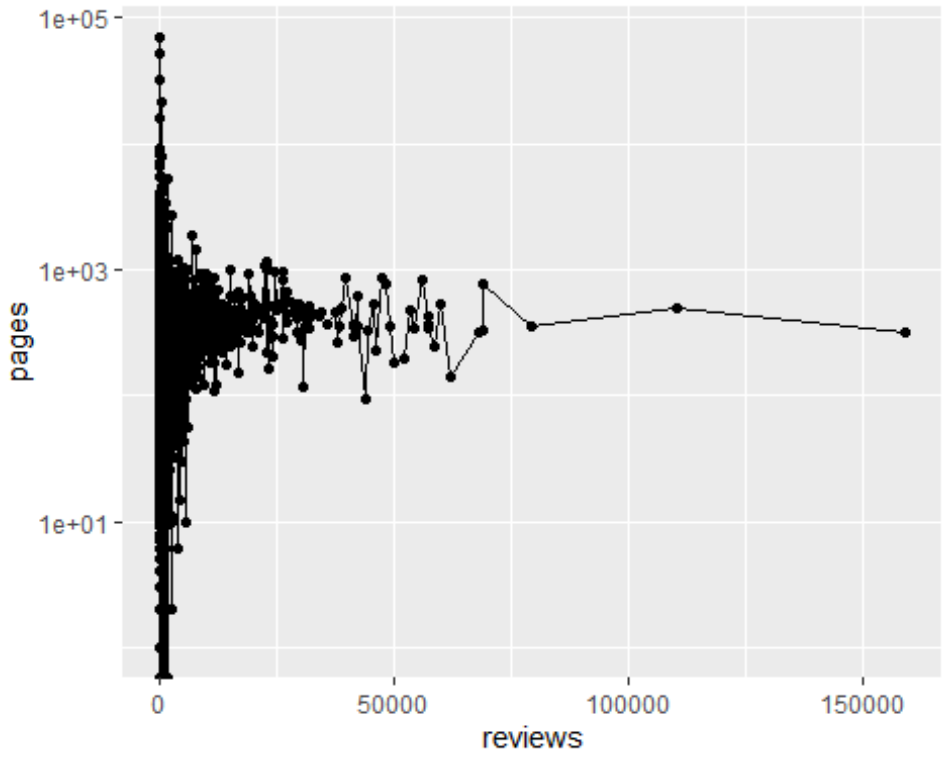


```
ggplot(data, aes(x = reviews, y = rating)) +
 geom_line() +
 geom_point()
```



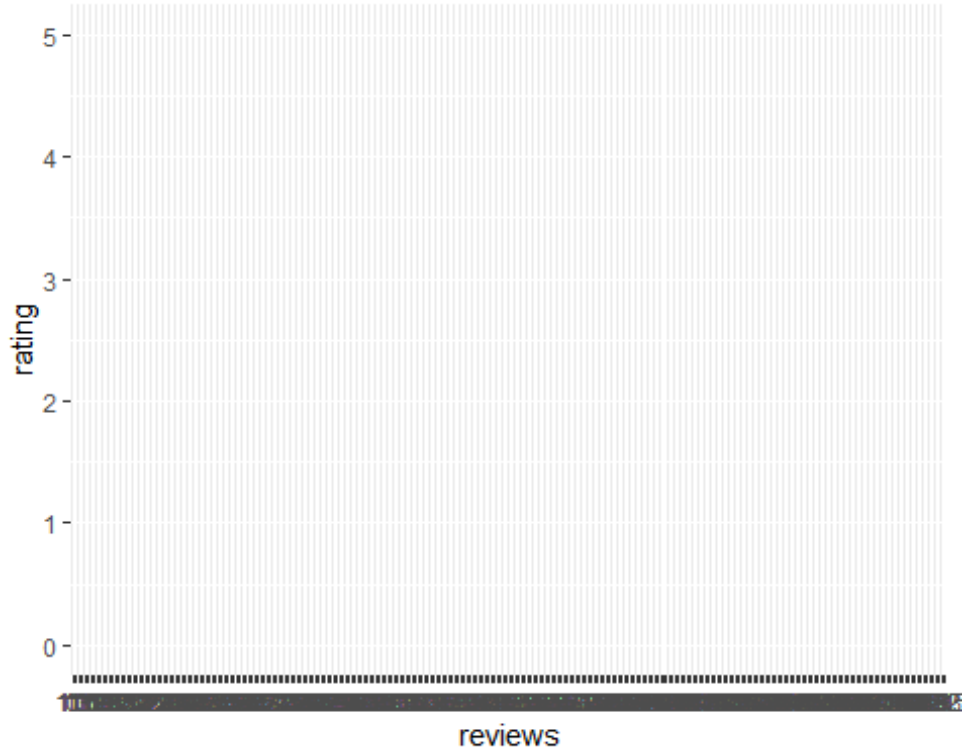

```
ggplot(data, aes(x = reviews, y = pages)) +
 geom_line() +
 geom_point() +
 scale_y_log10()
```

```
## Warning: Transformation introduced infinite values in continuous y-axis
## Transformation introduced infinite values in continuous y-axis
```

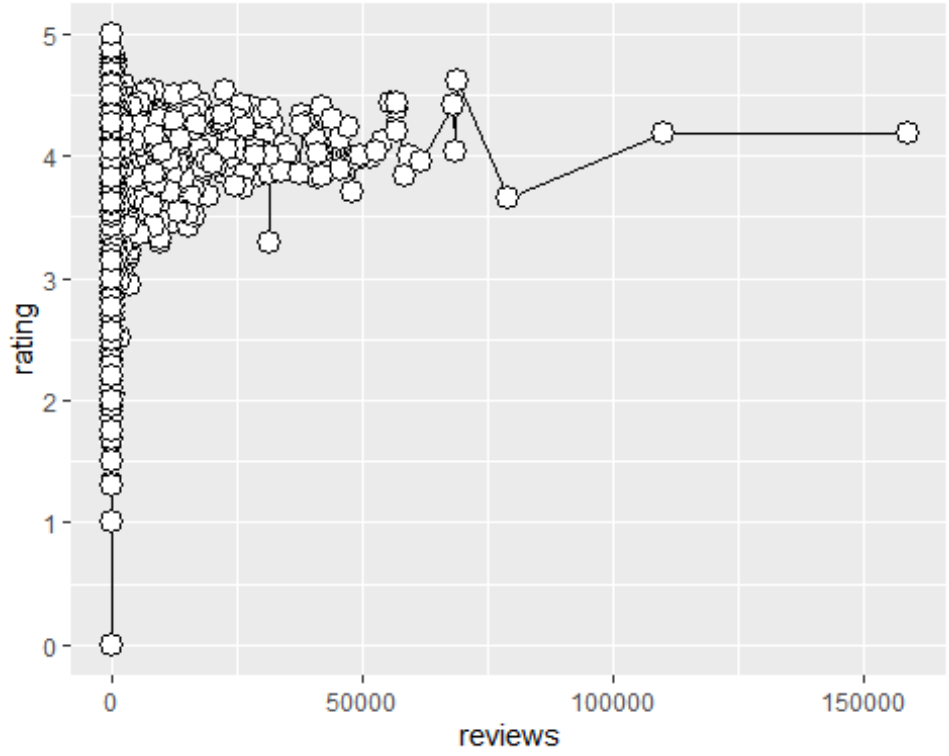



```
library(gcookbook)

ggplot(data1, aes(x =reviews, y = rating)) +
 geom_line()
```

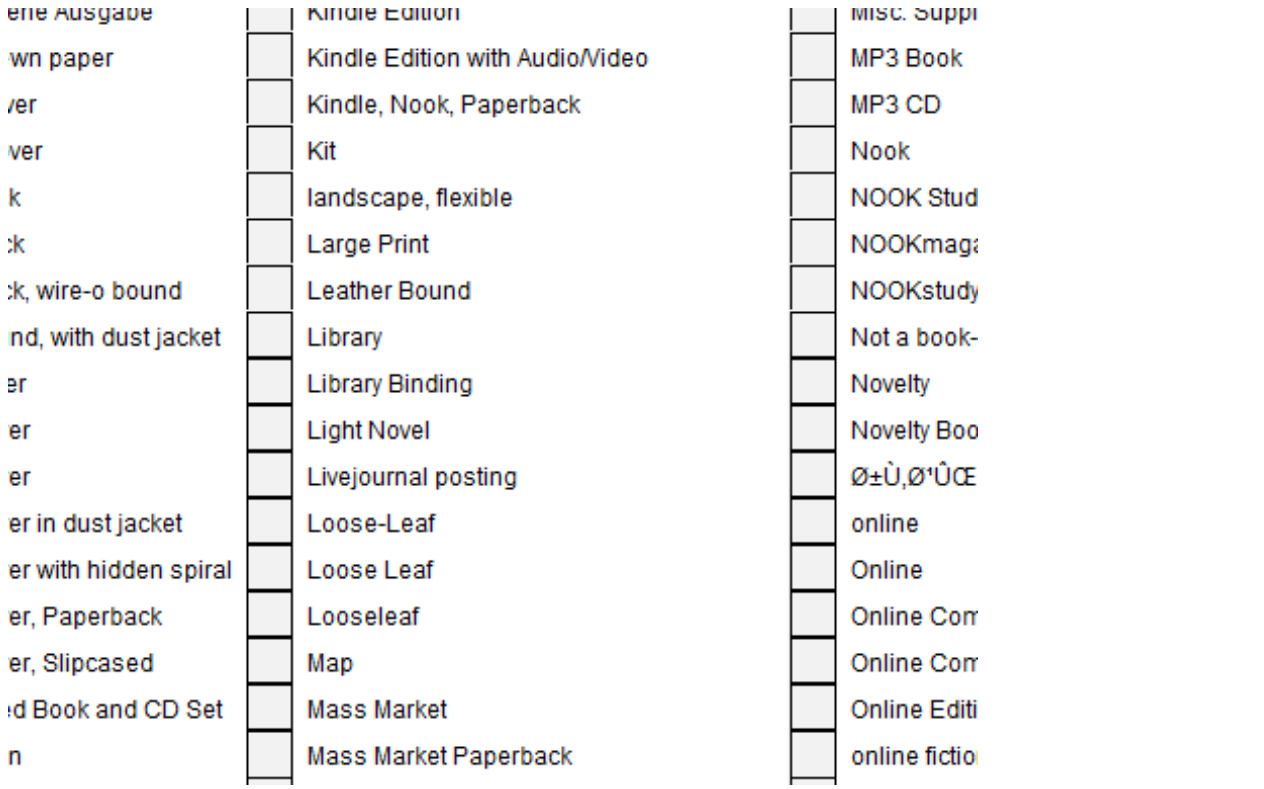



```
ggplot(data, aes(x = reviews, y = rating)) +
 geom_line() +
 geom_point(size = 4, shape = 21, fill = "white")
```

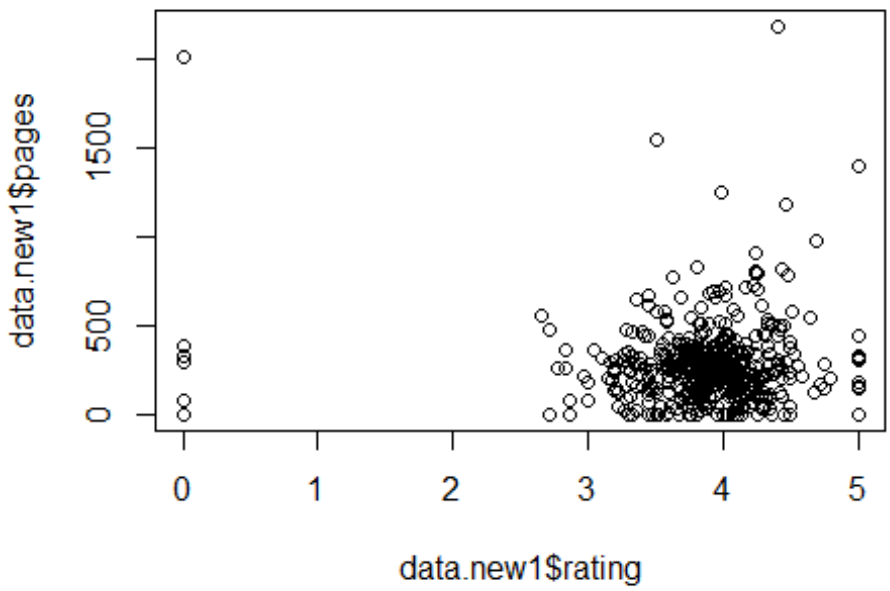


```
ggplot(data, aes(x =pages, y =reviews, fill =bookformat)) +
 geom_area(position = "fill", colour = "black", size = .2, alpha = .4) +
 scale_fill_brewer(palette = "Blues")
```

```
## Warning in RColorBrewer::brewer.pal(n, pal): n too large, allowed maximum
for palette Blues is 9
## Returning the palette you asked for with that many colors
```




```
data.new1<- data[1:500,]
plot(data.new1$rating,data.new1$pages)
```



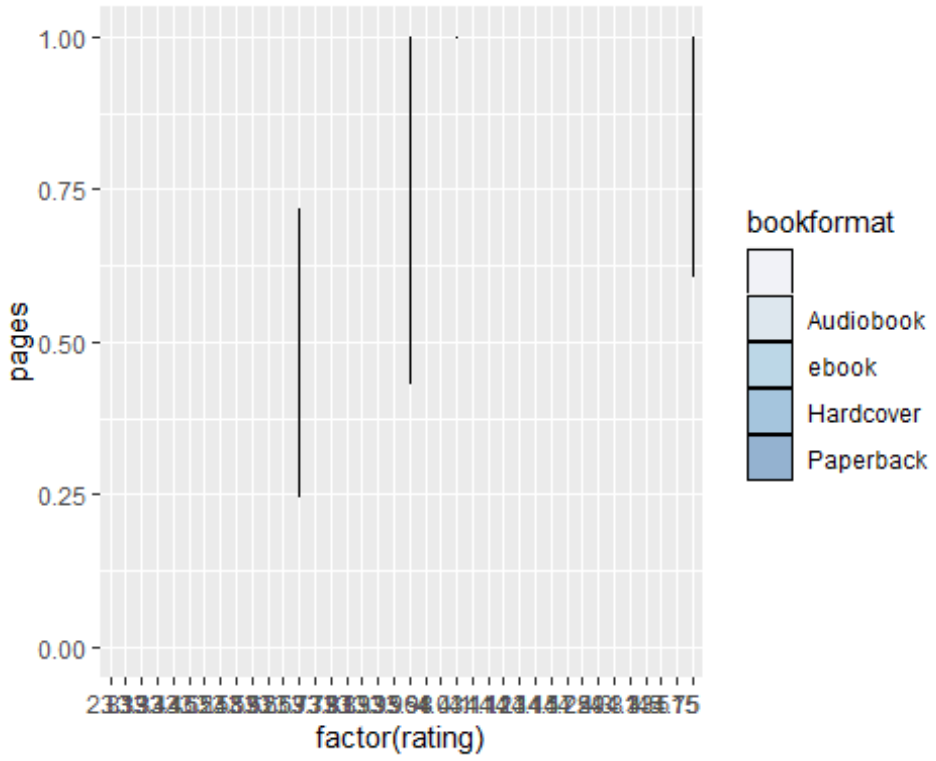

```
ggplot(data.new1[1:50,],aes(x=factor(rating),y =pages, fill =bookformat)) +
 geom_area(position = "fill", colour = "black", size = .2, alpha = .4) +
 scale_fill_brewer(palette = "Blues")
```

```
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; return
ing
## -Inf

## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; return
ing
## -Inf

## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; return
ing
## -Inf

## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; return
ing
## -Inf
```

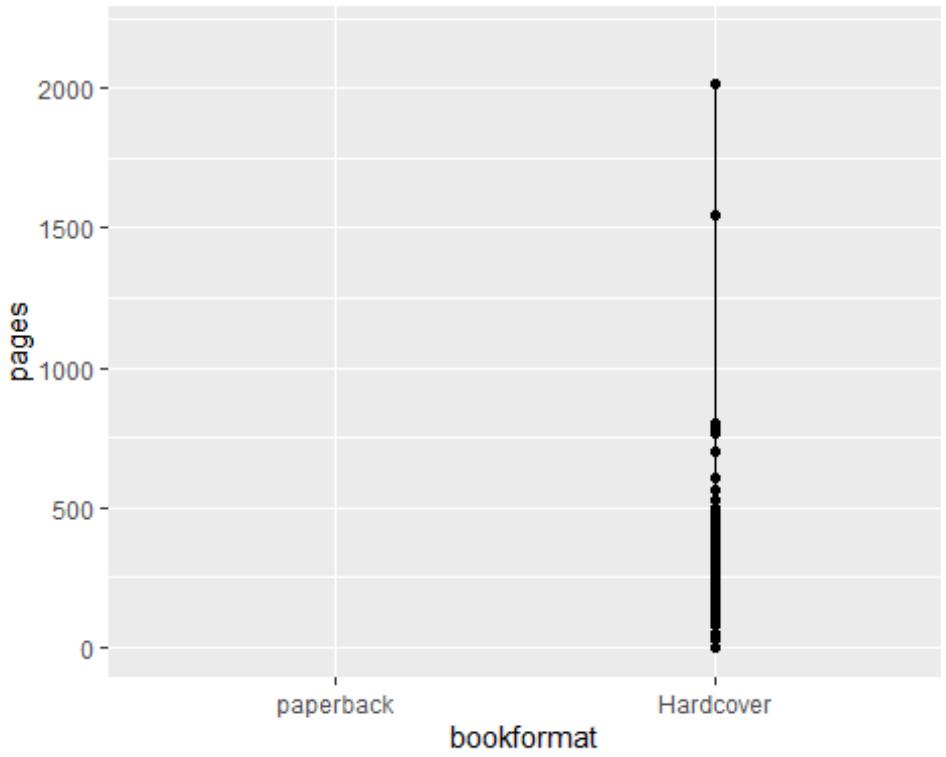


To change the order of items on a categrical axis and displaying subset of items on x axis using scale_x_discrete function

```
ggplot(data.new1, aes(x = bookformat, y = pages)) +
 geom_line() +
 geom_point() +
 scale_x_discrete(limits=c("paperback","Hardcover"))

## Warning: Removed 351 row(s) containing missing values (geom_path).

## Warning: Removed 351 rows containing missing values (geom_point).
```




```
ggplot(data.new1, aes(x = bookformat, y = pages)) + geom_line()+ scale_x_discrete(limits =rev(levels(data.new1$bookformat)))
```