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.

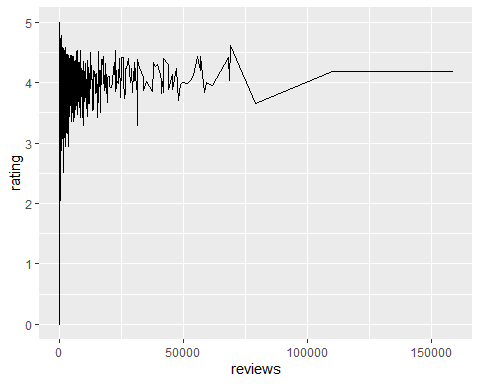
loading library to access various functions related to graph

library(ggplot2)

library(gcookbook)

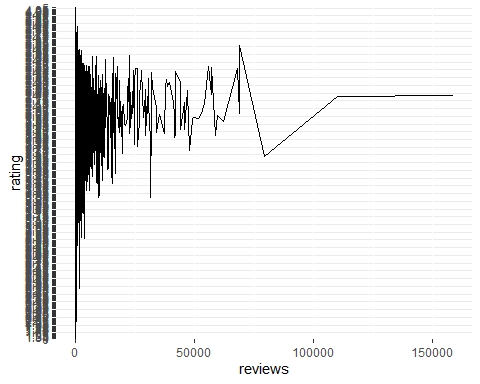
making a basic line graph.

ggplot(data, aes(x =reviews, y = rating)) +  
 geom\_line()

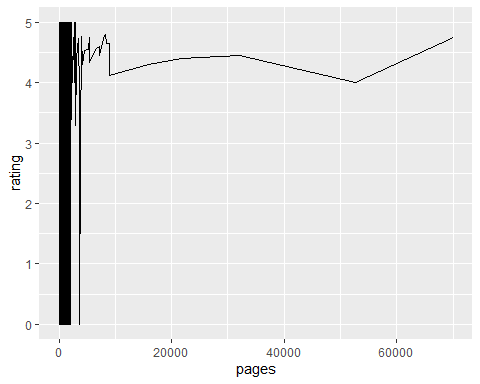
 Make a copy of the data

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

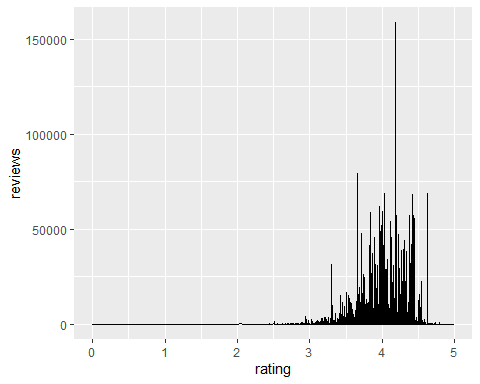
ggplot(data1, aes(x = reviews, y = rating, group = 1)) +  
 geom\_line()

 This will also give same result

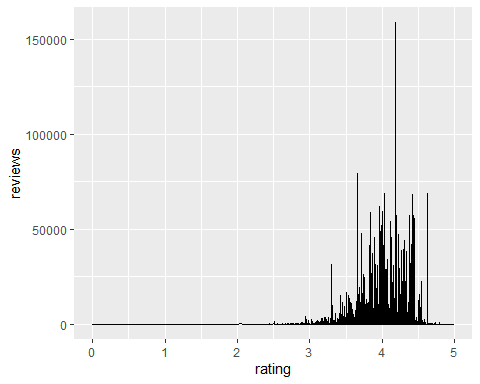
ggplot(data, aes(x = pages, y = rating)) +  
 geom\_line() +  
 ylim(0, max(data$rating))



ggplot(data, aes(x = rating, y = reviews)) +  
 geom\_line() +  
 expand\_limits(y = 0)

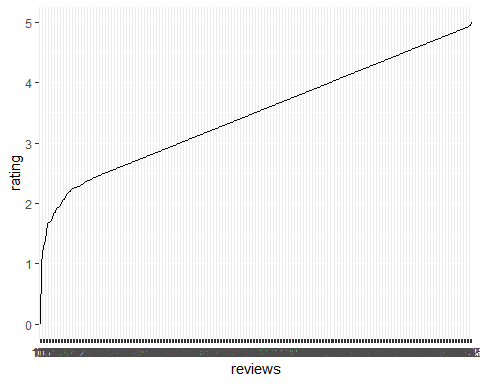


ggplot(data, aes(x =rating, y =reviews )) +  
geom\_line()

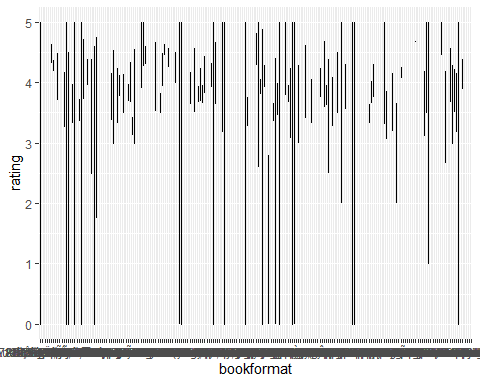


Make a copy of the data

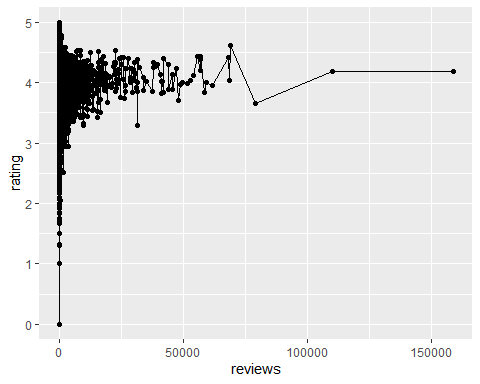
data1 <- data   
data1$reviews <- factor(data1$rating)  
ggplot(data1, aes(x = reviews, y = rating, group = 1)) +  
geom\_line()



ggplot(data, aes(x = bookformat, y = rating)) +  
 geom\_line() +  
 expand\_limits(y = 0)

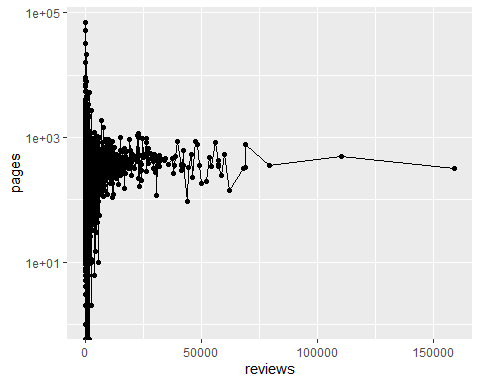


ggplot(data, aes(x = reviews, y = rating)) +  
 geom\_line() +  
 geom\_point()



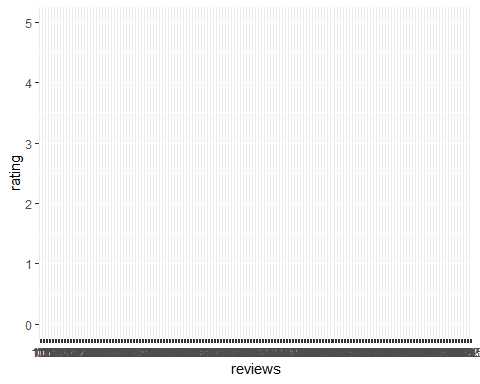
ggplot(data, aes(x = reviews, y = pages)) +  
 geom\_line() +  
 geom\_point() +  
 scale\_y\_log10()

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

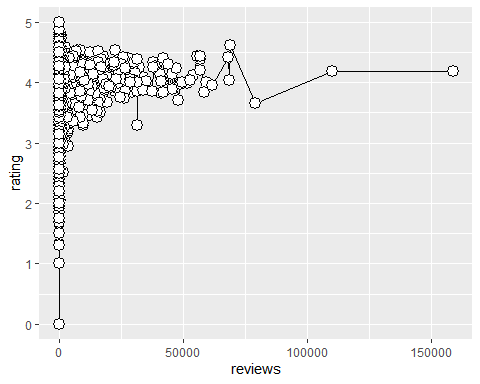


library(gcookbook)

ggplot(data1, aes(x =reviews, y = rating)) +  
 geom\_line()

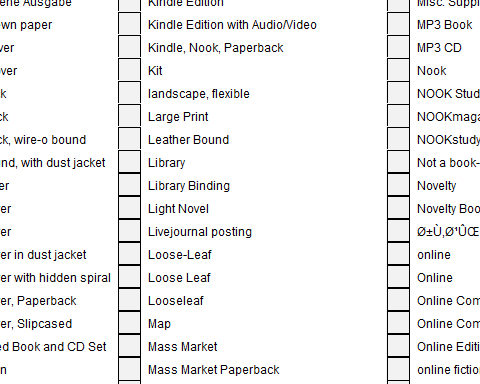


ggplot(data, aes(x = reviews, y = rating)) +  
 geom\_line() +  
 geom\_point(size = 4, shape = 21, fill = "white")

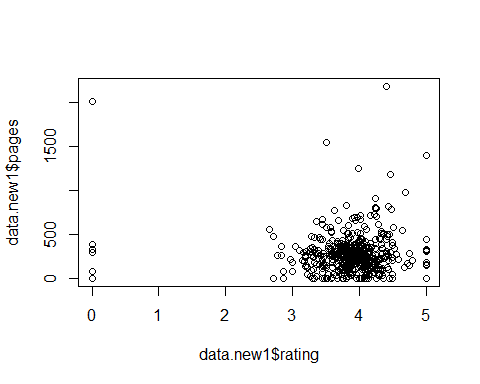


ggplot(data, aes(x =pages, y =reviews, fill =bookformat)) +  
 geom\_area(position = "fill", colour = "black", size = .2, alpha = .4) +  
 scale\_fill\_brewer(palette = "Blues")

## 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

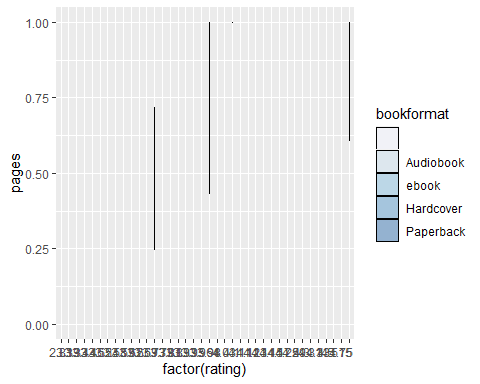


data.new1<- data[1:500,]  
plot(data.new1$rating,data.new1$pages)



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

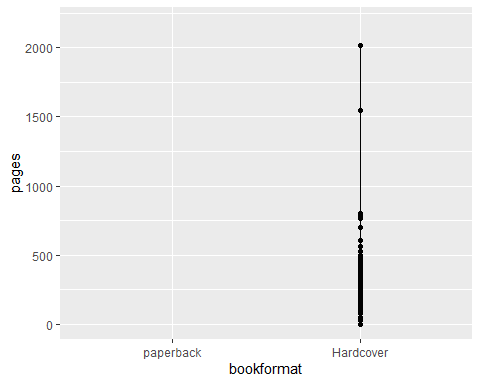


To change the order of items on a categrical axis and displaying subset of items on x axis using scale\_x\_discrete function

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)))

