Amazon Best Seller June 2021

Damaris Wanyoike

4/9/2022

# Data Description

The data set is about Amazon top 100 best selling products as of June 2021. The data is categorized into 8 columns;ASIN, Category, Product Link, No, of Sellers, Rating, Reviews Count and Price. The type of products in category include: Electronics, Video games, Camera & photo, Gift cards, Clothing, shoes & jewelry, Toys & games and Books. The data is stored in a csv format. Using the data set, we will hopefully answer the following questions. Which category of products is the most expensive? Which category of products has the most reviews? The rating averages for each product category.

#using read.csv to read data from a csv file  
AmazonBsel <- read.csv("Amazon\_Best\_Seller\_2021\_June.csv")

#Renaming the category column to product category  
rename(AmazonBsel, Productcategory = Category)

#Removing the dollar sign form price.  
AmazonBsel$Price = as.numeric(gsub("\\$","",AmazonBsel$Price))  
  
#Removing the commas from Reviews.Count  
AmazonBsel$Reviews.Count <- as.numeric(gsub(",","",AmazonBsel$Reviews.Count))

This data frame has 707 rows and 8 columns. The names of the columns and a brief description of each are in the table below;

text\_tbl<- data.frame(  
 Names =c("ASIN","Category", "Productlink", "No.of.Sellers", "Rating", "Reviews.Count", "Price"),  
 Description =c("Amazon Standard Identification Number", "Electronics,Video games,Camera & photo, Gift cards,Clothing,shoes & jewelry, Toys & games, Books","Link with respect to the Amazon product","Number of sellers selling the product", "Star value that a shopper submits for the product/features","Number of reviews submitted by customers for the product"," The price of the product")  
)  
text\_tbl

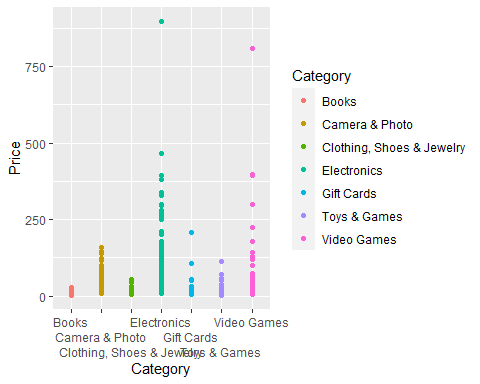
## Names  
## 1 ASIN  
## 2 Category  
## 3 Productlink  
## 4 No.of.Sellers  
## 5 Rating  
## 6 Reviews.Count  
## 7 Price  
## Description  
## 1 Amazon Standard Identification Number  
## 2 Electronics,Video games,Camera & photo, Gift cards,Clothing,shoes & jewelry, Toys & games, Books  
## 3 Link with respect to the Amazon product  
## 4 Number of sellers selling the product  
## 5 Star value that a shopper submits for the product/features  
## 6 Number of reviews submitted by customers for the product  
## 7 The price of the product

#picking three columns to use summary function  
Amazonsub<- select(AmazonBsel,Rating,Reviews.Count,Price)

#summary of the subset  
Summarytable <- summary(Amazonsub)  
Summarytable

## Rating Reviews.Count Price   
## Min. :1.400 Min. : 1 Min. : 0.88   
## 1st Qu.:4.500 1st Qu.: 5138 1st Qu.: 13.99   
## Median :4.700 Median : 18023 Median : 25.99   
## Mean :4.593 Mean : 77006 Mean : 55.69   
## 3rd Qu.:4.800 3rd Qu.: 49594 3rd Qu.: 50.00   
## Max. :5.000 Max. :854114 Max. :899.00

#The plot shows the price averages for each product category. It helps depicts which product category are the most and least expensive.  
library(ggplot2)  
ggplot(data = AmazonBsel) +  
 geom\_point(mapping = aes(x=Category, y=Price,color=Category)) +  
 scale\_x\_discrete(guide = guide\_axis(n.dodge = 3))



#The bar chart depicts the rating averages for each product category.  
ggplot(data = AmazonBsel) +  
 geom\_bar(mapping = aes(x= Category, y= Rating,fill=Category),stat = "identity") +  
 scale\_x\_discrete(guide = guide\_axis(n.dodge = 3))

