

Amazon_book

SMeher

2025-07-11

- AMAZON BOOK CASE STUDY #Package installation

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.2      v tibble    3.2.1
## v lubridate  1.9.4      v tidyr     1.3.1
## v purrr      1.0.4
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

-upload the .csv file

```
read_csv("Amazon_bestseller_book.csv")
```

```
## Rows: 550 Columns: 7
## -- Column specification -----
## Delimiter: ","
## chr (3): Name, Author, Genre
## dbl (4): User Rating, Reviews, Price, Year
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

## # A tibble: 550 x 7
##   Name                Author `User Rating` Reviews Price  Year Genre
##   <chr>              <chr>      <dbl>    <dbl> <dbl> <dbl> <chr>
## 1 10-Day Green Smoothie Cleanse JJ Sm~      4.7    17350      8  2016 Non ~
## 2 11/22/63: A Novel      Steph~      4.6     2052     22  2011 Fict~
## 3 12 Rules for Life: An Antidot~ Jorda~      4.7    18979     15  2018 Non ~
## 4 1984 (Signet Classics)  Georg~      4.7    21424      6  2017 Fict~
## 5 5,000 Awesome Facts (About Ev~ Natio~      4.8     7665     12  2019 Non ~
## 6 A Dance with Dragons (A Song ~ Georg~      4.4    12643     11  2011 Fict~
## 7 A Game of Thrones / A Clash o~ Georg~      4.7    19735     30  2014 Fict~
## 8 A Gentleman in Moscow: A Novel Amor ~      4.7    19699     15  2017 Fict~
## 9 A Higher Loyalty: Truth, Lies~ James~      4.7     5983      3  2018 Non ~
## 10 A Man Called Ove: A Novel   Fredr~      4.6    23848      8  2016 Fict~
## # i 540 more rows
```

-assign csv file to a vector named as amazon_book

```
amazon_book <- read.csv("Amazon_bestseller_book.csv")
```

View the first few rows

```
head(amazon_book)
```

```
##                                     Name
## 1                                10-Day Green Smoothie Cleanse
## 2                                11/22/63: A Novel
## 3                                12 Rules for Life: An Antidote to Chaos
## 4                                1984 (Signet Classics)
## 5 5,000 Awesome Facts (About Everything!) (National Geographic Kids)
## 6      A Dance with Dragons (A Song of Ice and Fire)
##      Author User.Rating Reviews Price Year      Genre
## 1      JJ Smith      4.7   17350     8 2016 Non Fiction
## 2      Stephen King      4.6    2052    22 2011      Fiction
## 3      Jordan B. Peterson      4.7   18979    15 2018 Non Fiction
## 4      George Orwell      4.7   21424     6 2017      Fiction
## 5 National Geographic Kids      4.8    7665    12 2019 Non Fiction
## 6      George R. R. Martin      4.4   12643    11 2011      Fiction
```

Get summary statistics

```
summary(amazon_book)
```

```
##      Name      Author      User.Rating      Reviews
## Length:550      Length:550      Min.   :3.300      Min.   : 37
## Class :character Class :character 1st Qu.:4.500      1st Qu.: 4058
## Mode  :character Mode  :character Median :4.700      Median : 8580
##                                     Mean  :4.618      Mean  :11953
##                                     3rd Qu.:4.800      3rd Qu.:17253
##                                     Max.   :4.900      Max.   :87841
##      Price      Year      Genre
## Min.   : 0.0      Min.   :2009      Length:550
## 1st Qu.: 7.0      1st Qu.:2011      Class :character
## Median :11.0      Median :2014      Mode  :character
## Mean   :13.1      Mean   :2014
## 3rd Qu.:16.0      3rd Qu.:2017
## Max.   :105.0      Max.   :2019
```

Check the structure of the data

```
str(amazon_book)
```

```
## 'data.frame': 550 obs. of 7 variables:
## $ Name : chr "10-Day Green Smoothie Cleanse" "11/22/63: A Novel" "12 Rules for Life: An Anti
## $ Author : chr "JJ Smith" "Stephen King" "Jordan B. Peterson" "George Orwell" ...
## $ User.Rating: num 4.7 4.6 4.7 4.7 4.8 4.4 4.7 4.7 4.7 4.6 ...
## $ Reviews : int 17350 2052 18979 21424 7665 12643 19735 19699 5983 23848 ...
## $ Price : int 8 22 15 6 12 11 30 15 3 8 ...
## $ Year : int 2016 2011 2018 2017 2019 2011 2014 2017 2018 2016 ...
```

```
## $ Genre      : chr  "Non Fiction" "Fiction" "Non Fiction" "Fiction" ...
```

Count books by genre

```
table(amazon_book$Genre)
```

```
##
##      Fiction Non Fiction
##      240      310
```

Count books by rating

```
table(amazon_book$User.Rating)
```

```
##
## 3.3 3.6 3.8 3.9  4 4.1 4.2 4.3 4.4 4.5 4.6 4.7 4.8 4.9
##   1   1   2   3 14   6   8 25  38  60 105 108 127 52
```

Average rating by genre

```
aggregate(`User.Rating` ~ Genre, data = amazon_book, mean)
```

```
##      Genre User.Rating
## 1      Fiction  4.648333
## 2 Non Fiction  4.595161
```

Most frequent authors

```
sort(table(amazon_book$Author), decreasing = TRUE) [1:10]
```

```
##
##      Jeff Kinney      Gary Chapman
##      12      11
##      Rick Riordan      Suzanne Collins
##      11      11
## American Psychological Association      Dr. Seuss
##      10      9
##      Gallup      Rob Elliott
##      9      8
##      Bill O'Reilly      Dav Pilkey
##      7      7
```

Top 5 frequent book name

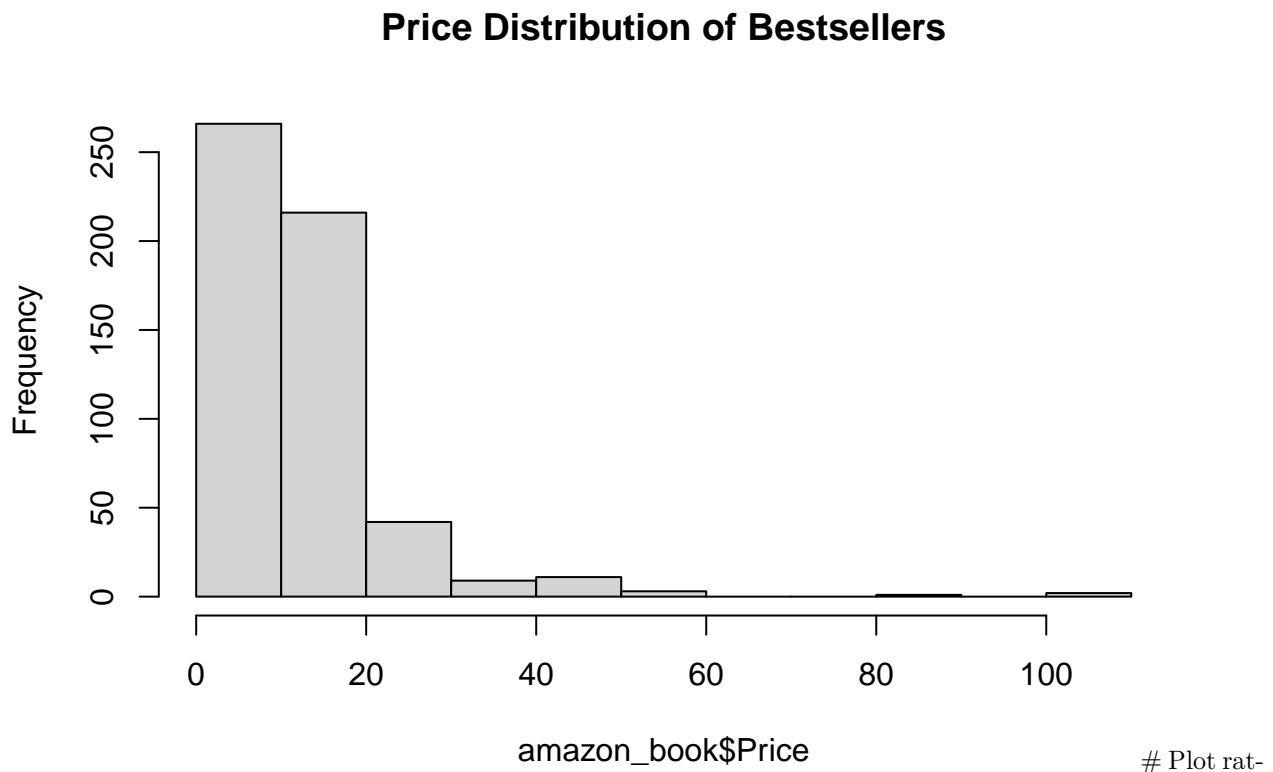
```
sort(table(amazon_book$Name), decreasing = TRUE) [1:5]
```

```
##
##      Publication Manual of the American Psychological Association, 6th Edition
##      10
##      StrengthsFinder 2.0
```

```
##
##                                Oh, the Places You'll Go!          9
##                                8
## The 7 Habits of Highly Effective People: Powerful Lessons in Personal Change
##                                7
##                                The Very Hungry Caterpillar
##                                7
##
```

Plot price distribution

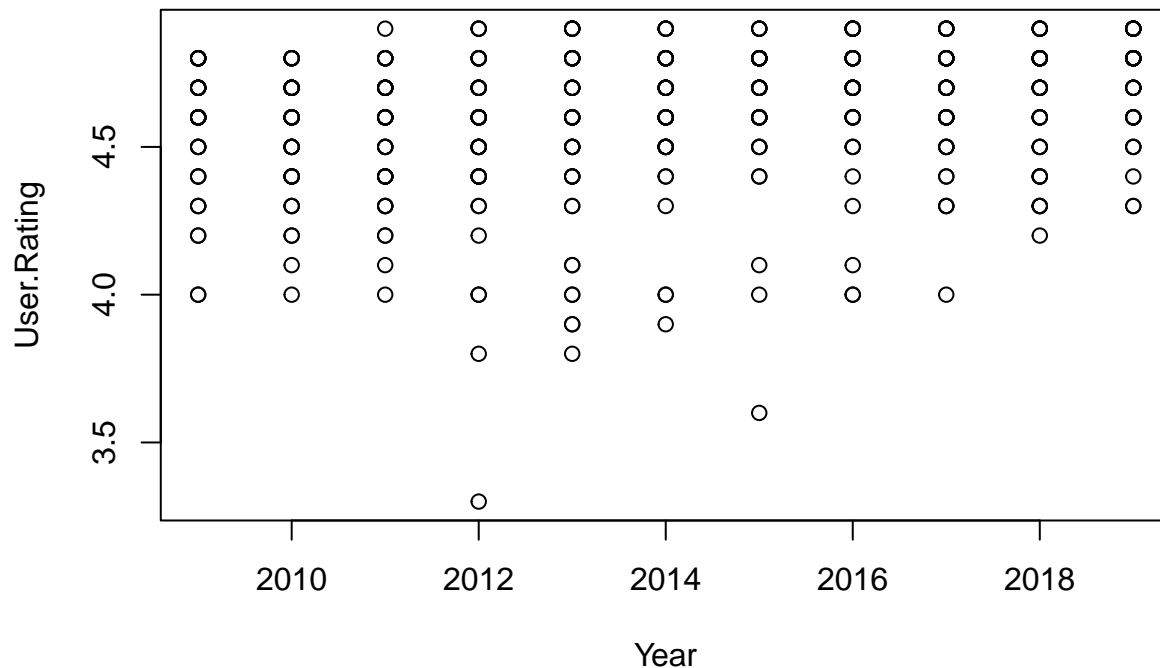
```
hist(amazon_book$Price, main = "Price Distribution of Bestsellers")
```



ings over time

Plot rat-

```
plot(`User.Rating` ~ Year, data = amazon_book)
```



Re-

name column to remove space

```
colnames(amazon_book)[3] <- "User_Rating"
view(amazon_book)
```

-Method 1: Using dplyr

```
library(dplyr)

top_10_expensive <- amazon_book %>%
  arrange(desc(Price)) %>%
  head(10) %>%
  select(Name, Author, Price, `User_Rating`, Reviews, Year, Genre)

# View the result
print(top_10_expensive)
```

```
##                                     Name
## 1 Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5
## 2 Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5
## 3                                     The Twilight Saga Collection
## 4                                     Hamilton: The Revolution
## 5                               The Book of Basketball: The NBA According to The Sports Guy
## 6                               Harry Potter Paperback Box Set (Books 1-7)
## 7 Publication Manual of the American Psychological Association, 6th Edition
## 8 Publication Manual of the American Psychological Association, 6th Edition
## 9 Publication Manual of the American Psychological Association, 6th Edition
## 10 Publication Manual of the American Psychological Association, 6th Edition
##                                     Author Price User_Rating Reviews Year
## 1   American Psychiatric Association   105      4.5      6679 2013
## 2   American Psychiatric Association   105      4.5      6679 2014
## 3         Stephenie Meyer              82      4.7      3801 2009
## 4       Lin-Manuel Miranda              54      4.9      5867 2016
## 5         Bill Simmons                  53      4.7        858 2009
```

```
## 6 J. K. Rowling 52 4.8 13471 2016
## 7 American Psychological Association 46 4.5 8580 2009
## 8 American Psychological Association 46 4.5 8580 2010
## 9 American Psychological Association 46 4.5 8580 2011
## 10 American Psychological Association 46 4.5 8580 2012
## Genre
## 1 Non Fiction
## 2 Non Fiction
## 3 Fiction
## 4 Non Fiction
## 5 Non Fiction
## 6 Fiction
## 7 Non Fiction
## 8 Non Fiction
## 9 Non Fiction
## 10 Non Fiction
```

-Method-2: Base R Approach # Top 10 highest price book # Sort the data by Price in descending order and extract top 10

```
top_10_expensive <- amazon_book[order(-amazon_book$Price), ][1:10, ]
```

Select relevant columns

```
top_10_expensive <- top_10_expensive[, c("Price", "Name", "Author", "User_Rating", "Reviews", "Year", "Genre")]
```

View the result

```
print(top_10_expensive)
```

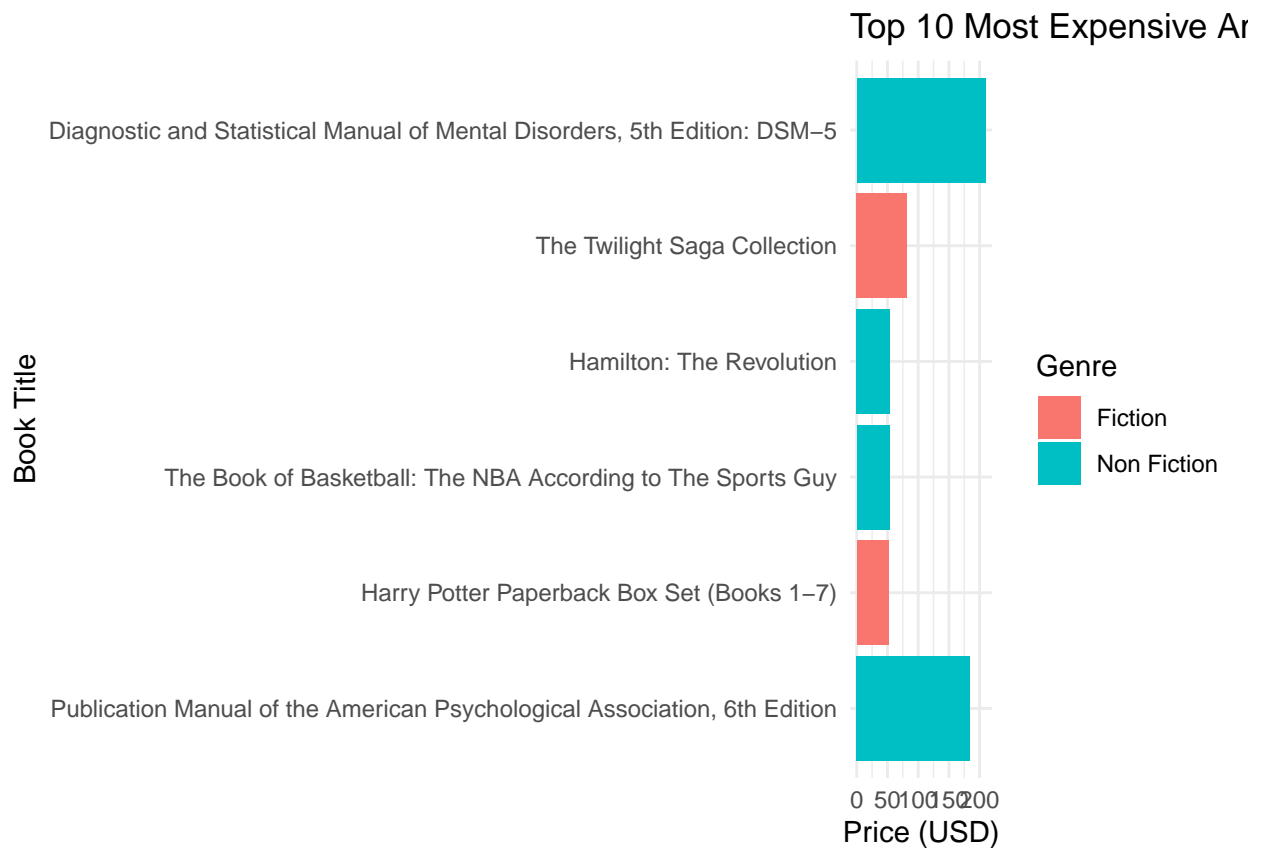
```
## Price
## 70 105
## 71 105
## 474 82
## 152 54
## 347 53
## 160 52
## 272 46
## 273 46
## 274 46
## 275 46
##
## Name
## 70 Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5
## 71 Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5
## 474 The Twilight Saga Collection
## 152 Hamilton: The Revolution
## 347 The Book of Basketball: The NBA According to The Sports Guy
## 160 Harry Potter Paperback Box Set (Books 1-7)
## 272 Publication Manual of the American Psychological Association, 6th Edition
## 273 Publication Manual of the American Psychological Association, 6th Edition
## 274 Publication Manual of the American Psychological Association, 6th Edition
## 275 Publication Manual of the American Psychological Association, 6th Edition
## Author User_Rating Reviews Year Genre
```

```
## 70    American Psychiatric Association      4.5    6679 2013 Non Fiction
## 71    American Psychiatric Association      4.5    6679 2014 Non Fiction
## 474          Stephenie Meyer                4.7    3801 2009    Fiction
## 152          Lin-Manuel Miranda              4.9    5867 2016 Non Fiction
## 347          Bill Simmons                   4.7     858 2009 Non Fiction
## 160          J. K. Rowling                  4.8   13471 2016    Fiction
## 272 American Psychological Association      4.5    8580 2009 Non Fiction
## 273 American Psychological Association      4.5    8580 2010 Non Fiction
## 274 American Psychological Association      4.5    8580 2011 Non Fiction
## 275 American Psychological Association      4.5    8580 2012 Non Fiction
```

–Visualization (Bar Plot of Top 10 Most Expensive Books)

```
library(ggplot2)

ggplot(top_10_expensive, aes(x = reorder(Name, Price), y = Price, fill = Genre)) +
  geom_bar(stat = "identity") +
  coord_flip() + # Horizontal bars for better readability
  labs(
    title = "Top 10 Most Expensive Amazon Bestseller Books",
    x = "Book Title",
    y = "Price (USD)"
  ) +
  theme_minimal()
```



–Top 10 highest-rated books #Method 1: Using dplyr

```
library(dplyr)
```

```
top_10_rated <- amazon_book %>%
  arrange(desc(`User_Rating`)) %>% # Sort by User Rating (descending)
  head(10) %>% # Select top 10
  select(Name, Author, `User_Rating`, Reviews, Price, Year, Genre) # Keep key columns

# View the result
print(top_10_rated)
```

```
##                                                                 Name
## 1                                                                 Brown Bear, Brown Bear, What Do You See?
## 2                                                                 Brown Bear, Brown Bear, What Do You See?
## 3                Dog Man and Cat Kid: From the Creator of Captain Underpants (Dog Man #4)
## 4    Dog Man: A Tale of Two Kitties: From the Creator of Captain Underpants (Dog Man #3)
## 5                Dog Man: Brawl of the Wild: From the Creator of Captain Underpants (Dog Man #6)
## 6                Dog Man: Brawl of the Wild: From the Creator of Captain Underpants (Dog Man #6)
## 7                Dog Man: Fetch-22: From the Creator of Captain Underpants (Dog Man #8)
## 8    Dog Man: For Whom the Ball Rolls: From the Creator of Captain Underpants (Dog Man #7)
## 9                Dog Man: Lord of the Fleas: From the Creator of Captain Underpants (Dog Man #5)
## 10 Goodnight, Goodnight Construction Site (Hardcover Books for Toddlers, Preschool Books for Kids)
##           Author User_Rating Reviews Price Year  Genre
## 1      Bill Martin Jr.      4.9   14344    5 2017 Fiction
## 2      Bill Martin Jr.      4.9   14344    5 2019 Fiction
## 3        Dav Pilkey      4.9    5062    6 2018 Fiction
## 4        Dav Pilkey      4.9    4786    8 2017 Fiction
## 5        Dav Pilkey      4.9    7235    4 2018 Fiction
## 6        Dav Pilkey      4.9    7235    4 2019 Fiction
## 7        Dav Pilkey      4.9   12619    8 2019 Fiction
## 8        Dav Pilkey      4.9    9089    8 2019 Fiction
## 9        Dav Pilkey      4.9    5470    6 2018 Fiction
## 10 Sherri Duskey Rinker      4.9    7038    7 2012 Fiction
```

-Method 2: Base R Approach # Sort by User Rating (descending) and take top 10 # Select relevant columns (optional) # View the result

```
top_10_rated <- amazon_book[order(-amazon_book$`User_Rating`), ][1:10, ]

top_10_rated <- top_10_rated[, c("Name", "Author", "User_Rating", "Reviews", "Price", "Year", "Genre")]

print(top_10_rated)
```

```
##                                                                 Name
## 41                                                                 Brown Bear, Brown Bear, What Do You See?
## 42                                                                 Brown Bear, Brown Bear, What Do You See?
## 82                Dog Man and Cat Kid: From the Creator of Captain Underpants (Dog Man #4)
## 83    Dog Man: A Tale of Two Kitties: From the Creator of Captain Underpants (Dog Man #3)
## 84                Dog Man: Brawl of the Wild: From the Creator of Captain Underpants (Dog Man #6)
## 85                Dog Man: Brawl of the Wild: From the Creator of Captain Underpants (Dog Man #6)
## 86                Dog Man: Fetch-22: From the Creator of Captain Underpants (Dog Man #8)
## 87    Dog Man: For Whom the Ball Rolls: From the Creator of Captain Underpants (Dog Man #7)
## 88                Dog Man: Lord of the Fleas: From the Creator of Captain Underpants (Dog Man #5)
## 147 Goodnight, Goodnight Construction Site (Hardcover Books for Toddlers, Preschool Books for Kids)
##           Author User_Rating Reviews Price Year  Genre
## 41      Bill Martin Jr.      4.9   14344    5 2017 Fiction
```



```
## 42      Bill Martin Jr.      4.9  14344    5 2019 Fiction
## 82      Dav Pilkey          4.9   5062    6 2018 Fiction
## 83      Dav Pilkey          4.9   4786    8 2017 Fiction
## 84      Dav Pilkey          4.9   7235    4 2018 Fiction
## 85      Dav Pilkey          4.9   7235    4 2019 Fiction
## 86      Dav Pilkey          4.9  12619    8 2019 Fiction
## 87      Dav Pilkey          4.9   9089    8 2019 Fiction
## 88      Dav Pilkey          4.9   5470    6 2018 Fiction
## 147 Sherri Duskey Rinker    4.9   7038    7 2012 Fiction
```

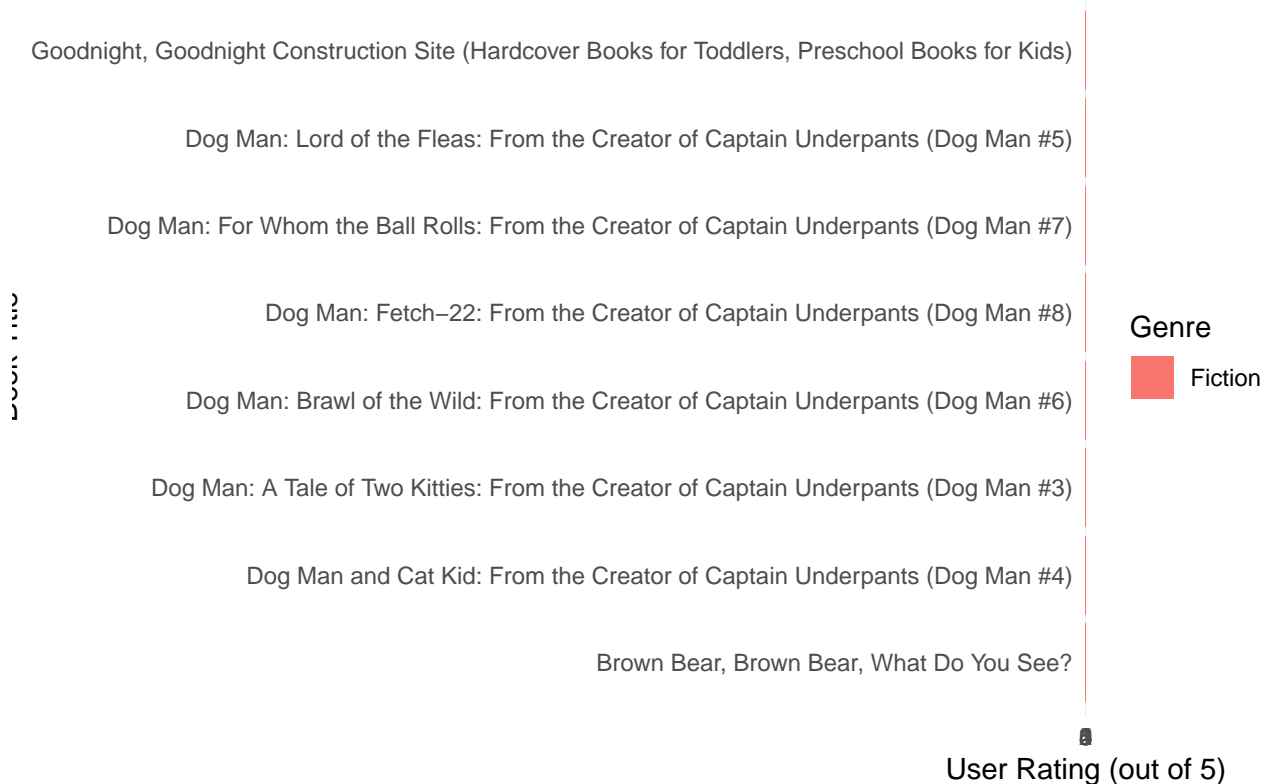
– Visualization (Top 10 Rated Books)

```
library(ggplot2)

ggplot(top_10_rated, aes(x = reorder(Name, `User_Rating`), y = `User_Rating`, fill = Genre)) +
  geom_bar(stat = "identity") +
  coord_flip() + # Horizontal bars for readability
  labs(
    title = "Top 10 Highest-Rated Amazon Bestseller Books",
    x = "Book Title",
    y = "User Rating (out of 5)"
  ) +
  theme_minimal() +
  scale_y_continuous(limits = c(0, 5)) # Adjust if ratings exceed 5
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## (`geom_bar()`).
```

Top 10 High



Key Findings from the Amazon Bestseller Books Case Study

Genre Distribution

Non-Fiction (310 books) slightly outnumbers Fiction (240 books) in the dataset, suggesting a strong market for informative/self-help content.

Pricing Insights

The most expensive book (*Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5*) costs \$105, while the average price is \$13.1.

Top 10 expensive books are dominated by Non-Fiction (7 out of 10), including academic/reference books.

User Ratings

Average rating: 4.62/5, with Fiction books having a slightly higher average (4.65) than Non-Fiction (4.60).

Top-rated books (4.9/5) are primarily children's books (e.g., *Brown Bear, Brown Bear, What Do You See?*) and *Dog Man* series.

Trends Over Time

Higher-rated books tend to have lower prices (e.g., top-rated books average \$6.2 vs. dataset average of \$13.1).

Books with 5,000–20,000 reviews typically have ratings of 4.7+, indicating popularity correlates with quality perception.

Author and Title Patterns

Most frequent authors: Jeff Kinney (12 books), Gary Chapman (11), and Rick Riordan (11).

Top recurring titles: *Publication Manual of the American Psychological Association* (10 editions) and *StrengthsFinder 2.0* (9 editions).

Recommendations

For Publishers & Authors

Leverage Non-Fiction Demand: Invest in high-quality reference/self-help books, which dominate the high-price segment.

Children's Books: Expand series like *Dog Man*, as they achieve both high ratings and moderate pricing.

For Pricing Strategy

Competitive Pricing: Top-rated books are often priced below \$10; avoid overpricing to maintain ratings.

Bundle Offers: For expensive titles (e.g., *DSM-5*), offer discounts or digital versions to increase accessibility.

For Amazon Sellers

Promote High-Rated, Affordable Books: Use algorithms to highlight books with 4.7+ ratings and <\$15 prices, as they likely drive customer satisfaction.

Genre-Based Campaigns: Target Fiction readers with highly rated novels and Non-Fiction readers with practical guides.

Future Research

Analyze profit margins (price vs. production cost) to identify the most lucrative genres.

Study seasonal trends (e.g., academic books selling more in fall).

Conclusion

The analysis reveals that high ratings don't require high prices, as evidenced by top-performing children's books. Non-Fiction dominates in volume and premium pricing, while Fiction excels in user satisfaction. Strategic pricing, genre focus, and series expansion are key to leveraging Amazon's bestseller trends.

Final Takeaway:

Balance quality (ratings), affordability, and genre to maximize sales and customer loyalty in the competitive bookselling market.