# RWorksheet#5_Group(Basa,Llanera,Tuares)

## 2024-11-06

```r
# Load necessary libraries
library(polite)
library(rvest)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(purrr)
library(ggplot2)


scrape_amazon_category <- function(category_url, category_name) {

  session <- bow(category_url, user_agent = "Educational")


  page <- scrape(session)


  products <- page %>%
    html_nodes(".s-main-slot .s-result-item") %>%
    map_df(~ {
      tibble(
        Title = .x %>% html_node("h2") %>% html_text(trim = TRUE),
        Price = .x %>% html_node(".a-price .a-offscreen") %>% html_text(trim = TRUE),
        Description = .x %>% html_node(".a-text-normal") %>% html_text(trim = TRUE),
        Rating = .x %>% html_node(".a-icon-alt") %>% html_text(trim = TRUE),
        Reviews = .x %>% html_node(".a-size-small .a-link-normal") %>% html_text(trim = TRUE),
        Category = category_name  # Add category name here
      )
    })

  return(products)
}


categories <- list(
 computers = 'https://www.amazon.com/s?k=computers',
```

```
    luggage = 'https://www.amazon.com/s?k=luggage',
    dress = 'https://www.amazon.com/s?k=dress&ref=nav_bb_sb',
    TV = 'https://www.amazon.com/s?k=TV',
    ballpen = 'https://www.amazon.com/s?k=ballpen'
)


all_products <- tibble()


for (category_name in names(categories)) {
  category_url <- categories[[category_name]]
  category_products <- scrape_amazon_category(category_url, category_name)
  all_products <- bind_rows(all_products, category_products)


  if (nrow(all_products) > 30) {
    all_products <- all_products %>% slice(1:30)
  }
}


all_products$Price <- as.numeric(gsub("\\$", "", gsub(",", "", all_products$Price)))
all_products$Rating <- as.numeric(gsub(" out of 5 stars", "", all_products$Rating))
```

```
## Warning: NAs introduced by coercion
```

```
print(all_products)
```

```
## # A tibble: 30 x 6
##    Title                                Price Description Rating Reviews Category
##    <chr>                                <dbl> <chr>       <dbl> <chr>   <chr>
## 1  <NA>                                 NA    <NA>        NA    <NA>    compute~
## 2  "Results"                            NA    "Check eac~ NA    <NA>    compute~
## 3  "HP Chromebook 14 inch Laptop, FHD~  160.  "HP Chrome~ 4.3   70      compute~
## 4  "HP 15.6\" Portable Laptop (Includ~  277   "HP 15.6\"~ 4.1   5 capa~ compute~
## 5  "Acer Aspire 3 A315-24P-R7VH Slim ~  280.  "Acer Aspi~ 4.4   3 capa~ compute~
## 6  "More results"                       NA    <NA>        NA    <NA>    compute~
## 7  "HP Newest 14\" Ultral Light Lapto~  250.  "HP Newest~ 4     1,959   compute~
## 8  "HP Portable Laptop, Student and B~  210   "HP Portab~ 4.1   5 capa~ compute~
## 9  "HP 14 Laptop, Intel Celeron N4020~  183   "HP 14 Lap~ 4     1,881   compute~
## 10 "HP 2022 Newest All-in-One Desktop~  425.  "HP 2022 N~ 4.3   951     compute~
## # i 20 more rows
```

```
ggplot(all_products, aes(x = Title, y = Price, fill = Category)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  theme_minimal() +
  labs(title = "Price of Products by Category", x = "Product", y = "Price ($)")
```

```
## Warning: Removed 9 rows containing missing values or values outside the scale range
## (`geom_bar()`).
```

N4120, 16GB RAM, 288GB Storage (128GB eMMC + 160GB Docking Station Set), Intel UHD Graphics, 720p Webcam,
usiness, 14" HD Display, Intel Quad–Core N4120, 16GB DDR4 RAM, 64GB eMMC, 1 Year Office 365, Webcam, RJ–4!
omputer, 21.5" FHD Display, Intel Celeron J4025 Processor, 16GB RAM, 256GB SSD, Webcam, HDMI, RJ–45, Wired I
Business, Intel Quad–Core N4120, 8GB RAM, 192GB Storage(64GB eMMC+128GB Micro SD), 1 Year Office 365, W
) G1 SFF Desktop Computer Package – Intel Core i5 3.3GHz, 32GB RAM, New 1TB SSD, KOORUI 24–inch Monitor, /
ktop PC Computer Intel Core i5 3.1–GHz, 8 gb Ram, 1 TB Hard Drive, DVDRW, 19 Inch LCD Monitor, Keyboard, Mous
HP Chromebook 14 inch Laptop, FHD Display, Intel N100, 4 GB RAM, 64 GB eMMC, Intel UHD G
FHD Display, Intel Celeron J4025 Processor, 16GB RAM, 512GB PCIe SSD, Webcam, HDMI, RJ–45, Wired Keyboard
HP 17 Business Laptop, 17.3" HD+ Display, 11th Gen Intel Core i3–1125G4 Processor, 32GB RAM, 1TB SSD, Wi–F
icrosoft 365), HD Display, Intel Quad–Core N200 Processor, 16GB RAM, 128GB Storage, Wi–Fi 6, Webcam, HDMI, N
64 GB Storage, 14–inch Micro–edge HD Display, Windows 11 Home, Thin & Portable, 4K Graphics, One Year of Micro
Gen Intel 4–Core Processor | 16GB DDR4 RAM | 256GB SSD + 1TB HDD | Intel UHD Graphics 730 | WiFi 6 | DVD–R
0 SFF Desktop PC Intel i7–7700 4–Cores 3.60GHz 32GB DDR4 1TB SSD WiFi BT HDMI Duel Monitor Support Wind

16:10 FHD 165Hz Display, NVIDIA® GeForce RTXTM 4060, Intel Core i7–13650HX, 16GB DDR5, 1TB PCIe Gen4 SS
CM14 Laptop, 14" HD Anti–Glare Display (1366x768), MediaTek Kompanio 520, 4GB RAM, 64GB eMMC, ChromeOS,

PS Display | AMD Ryzen 3 7320U Quad–Core Processor | AMD Radeon Graphics | 8GB LPDDR5 | 128GB NVMe SSI

```r
ggplot(all_products, aes(x = Title, y = Rating, fill = Category)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  theme_minimal() +
  labs(title = "Ratings of Products by Category", x = "Product", y = "Rating (out of 5)")
```

```
## Warning: Removed 9 rows containing missing values or values outside the scale range
## (`geom_bar()`).
```

```r
ranked_by_price <- all_products %>%
  group_by(Category) %>%
  arrange(Price) %>%
  mutate(Price_Rank = row_number())


ranked_by_rating <- all_products %>%
  group_by(Category) %>%
  arrange(desc(Rating)) %>%
  mutate(Rating_Rank = row_number())
```
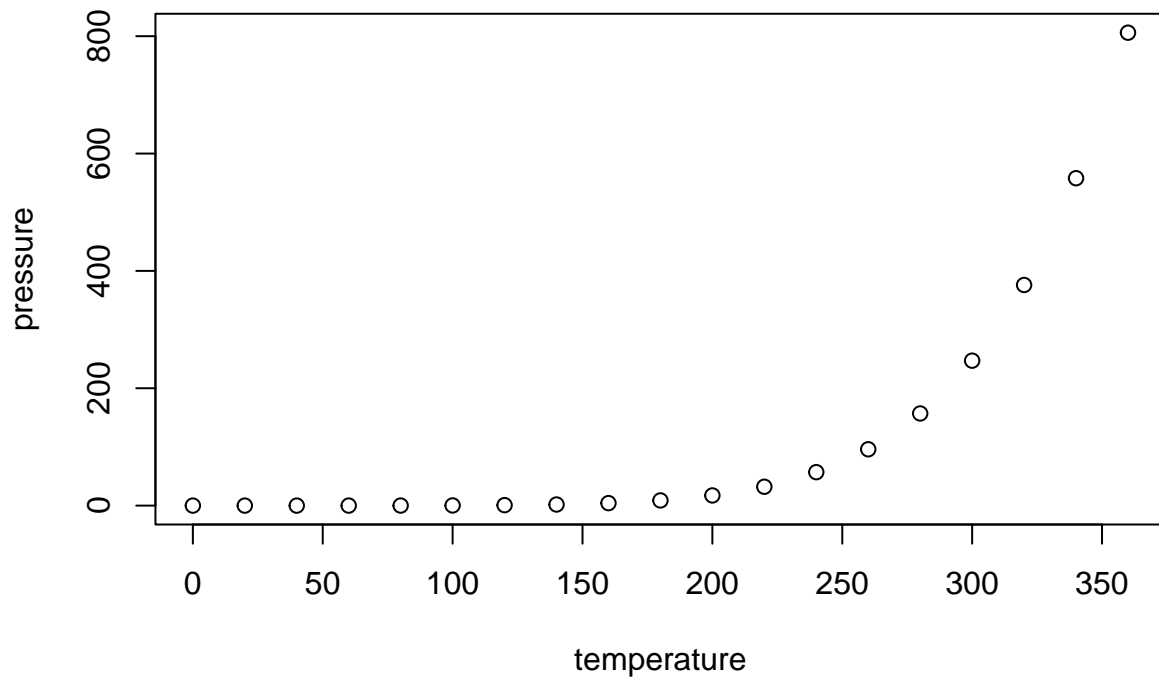
```r
print(ranked_by_price)
```

```
## # A tibble: 30 x 7
## # Groups:   Category [2]
##    Title                  Price Description Rating Reviews Category Price_Rank
##    <chr>                  <dbl> <chr>        <dbl> <chr>   <chr>         <int>
##  1 "Amazon Basics"         74.5 "Expandabl~    4.5 5 sizes luggage          1
##  2 "Travelers Club"        95.2 "Midtown H~    4.1 19,400  luggage          2
##  3 "HP Elite Desktop PC Co~ 123  "HP Elite ~    3.8 7,495   compute~         1
##  4 "Cosbarn"              130.  "Luggage S~    4.3 91      luggage          3
##  5 "Samsonite"            149  "Omni PC H~    4.4 6 sizes luggage          4
##  6 "Samsonite"            158.  "Evolve Se~    4.4 8 sizes luggage          5
##  7 "HP Chromebook 14 inch ~ 160.  "HP Chrome~    4.3 70      compute~         2
##  8 "ASUS Chromebook CM14 L~ 169   "ASUS Chro~    4.2 1,095   compute~         3
##  9 "HP 14 Laptop, Intel Ce~ 183   "HP 14 Lap~    4   1,881   compute~         4
## 10 "HP Portable Laptop, St~ 210   "HP Portab~    4.1 5 capa~ compute~         5
## # i 20 more rows
```

```r
print(ranked_by_rating)
```

```
## # A tibble: 30 x 7
## # Groups:   Category [2]
##    Title                  Price Description Rating Reviews Category Rating_Rank
##    <chr>                  <dbl> <chr>        <dbl> <chr>   <chr>          <int>
##  1 "Amazon Basics"         74.5 "Expandabl~    4.5 5 sizes luggage           1
##  2 "Acer Aspire 3 A315-2~  280.  "Acer Aspi~    4.4 3 capa~ compute~          1
##  3 "HP Newest All-in-One~  400.  "HP Newest~    4.4 8 sizes compute~          2
##  4 "Samsonite"            158.  "Evolve Se~    4.4 8 sizes luggage           2
##  5 "Samsonite"            149  "Omni PC H~    4.4 6 sizes luggage           3
##  6 "HP Chromebook 14 inc~  160.  "HP Chrome~    4.3 70      compute~          3
##  7 "HP 2022 Newest All-i~  425.  "HP 2022 N~    4.3 951     compute~          4
##  8 "Dell Vostro 3910 Des~  500.  "Dell Vost~    4.3 88      compute~          5
##  9 "HP Stream 14\" HD Br~  309   "HP Stream~    4.3 4 capa~ compute~          6
## 10 "ASUS ROG Strix G16 (~ 1270.  "ASUS ROG ~    4.3 1,421   compute~          7
## # i 20 more rows
```

## 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.