Solutions for Homework I

Uğurcan Bayraktar

11/15/2021

Question 1

Firstly, we import the tidyverse.

```
library(tidyverse)
options(dplyr.print_max = 1e9)
options(scipen = 100)
```

Then, the data should be read. It is assigned to **revenue_data.**

```
revenue_data <- read_csv("data/Revenue_Chart_Full_Data_data.csv")
```

Question 2

Year (copy), *Format*, *Value (Actual)* columns are renamed below.

```
revenue_annually = revenue_data %>%
                     select(`Year (copy)`, Format, `Value (Actual)`) %>%
                     rename(Year = `Year (copy)`,
                            Format = Format,
                            Value = `Value (Actual)`)
```

This is the first 6 rows of the subset.

Year	Format	Value
2005	Cassette	13.100000
2015	CD Single	1.196947
2015	Paid Subscription	1156.708514
2017	Download Single	667.875936
1986	Vinyl Single	228.100000
2003	DVD Audio	8.000000

Then the year column is sorted.

```
revenue_annually = revenue_annually %>%
                   filter(Year >= 2018) %>%
                   arrange(Year)
```

Year	Format	Value
2018	Vinyl Single	5.7206009
2018	SoundExchange Distributions	952.8000000
2018	Ringtones & Ringbacks	24.9578598
2018	SACD	0.8596997

Year	Format	Value
2018	Download Music Video	2.2208106
2018	Paid Subscription	4614.0185024

Finally, annually revenue is calculated from 2018 to 2020.

```
revenue_annually = revenue_annually %>%
                     group_by(Year) %>%
                     summarise(`Value (Yearly)` = sum(Value))
```

Year	Value (Yearly)
2018	9738.197
2019	11130.353
2020	12153.351

Question 3

Created a subset that includes the streaming formats.

```
streaming_revenue = revenue_data %>%
                      select(Year, Streaming = Format, Value = `Value (Actual)`) %>%
                      filter(Streaming %in% c("Paid Subscription",
                                              "On-Demand Streaming (Ad-Supported)",
                                              "Other Ad-Supported Streaming",
                                              "SoundExchange Distributions",
                                              "Limited Tier Paid Subscription"),
                                               Year == 2020) %>%
                      arrange(Streaming)
```

Year	Streaming	Value
2020	Limited Tier Paid Subscription	723.6147
2020	On-Demand Streaming (Ad-Supported)	1183.1214
2020	Other Ad-Supported Streaming	211.2476
2020	Paid Subscription	7009.1655
2020	SoundExchange Distributions	947.4000

The total streaming music revenue is approximately $10.1 billion in 2020.

```
streaming_revenue = streaming_revenue %>%
                      group_by(Year)%>%
                      summarise(`Streaming Revenue` = sum(Value))
```

Year	Streaming Revenue
2020	10074.55

Question 4

To begin with, I created a subset named as **revenue_share**, then made a new column which is named as **Categories** consists of *Streaming*, *Synchronization*, *Physical*, *Digital Downloads* and *Other* categories.

```
revenue_share = revenue_data %>%
                  select(Year, Format, Value= `Value (Actual)`) %>%
                  filter(Year == 2020) %>%
                  mutate(Category = case_when(
                    Format %in% c("Paid Subscription",
                                  "On-Demand Streaming (Ad-Supported)",
                                  "Other Ad-Supported Streaming",
                                  "SoundExchange Distributions",
                                  "Limited Tier Paid Subscription") ~ "Streaming",
                    Format %in% c("Synchronization") ~ "Synchronization",
                    Format %in% c("LP/EP", "CD") ~ "Physical",
                    Format %in% c("Download Album",
                                  "Download Single") ~ "Digital Downloads",
                    TRUE ~ "Other"))
```

Finally, I created a new column includes a revenue shares as percentages and sorted in decreasing order.

```
revenue_share = revenue_share %>%
                  group_by(Category) %>%
                  summarise(Revenue_Total = sum(Value)) %>%
                  mutate(Share = round(Revenue_Total / sum(Revenue_Total),2)) %>%
                  arrange(-Share)
```

Category	Revenue_Total	Share
Streaming	10074.5491	0.83
Physical	1102.9506	0.09
Digital Downloads	632.3201	0.05
Synchronization	265.2361	0.02
Other	78.2956	0.01

Question 5

Radio revenue data is created below. It includes *SoundExchange Distributions* and *Other Ad-Supported Streaming* revenues from 2018 to 2020.

```
radio_revenue <- revenue_data %>%
                    select(Year, Format, Value = `Value (For Charting)`) %>%
                    filter(Year >= 2018,
                           Format %in% c("SoundExchange Distributions",
                                         "Other Ad-Supported Streaming")) %>%
                    arrange(Year)
```

Here is another subset created for charting the **annually revenues** in the **radio revenue** data.

```
radrev_annually <- radio_revenue %>%
                    group_by(Year) %>%
                    summarise(Value_Annually = sum(Value))
```

Graph is created below.

```
#Creating graph with ggplot
#Filled the stack bars in order with levels
g <- ggplot(radio_revenue, aes(x = Year,
                               y = Value,
                               fill = factor(Format,
```

```
                                          levels = c("Other Ad-Supported Streaming",
                                                     "SoundExchange Distributions")))) +
      geom_bar(stat = "identity") + #Bar graph will be used

      #Labels are created as same as the original figure.
      labs(x = "", y = "$ MILLIONS",
           title = "U.S. DIGITAL AND CUSTOMIZED RADIO REVENUES",
           subtitle = "Source: RIAA")+

      #Theme is updated: title, subtitle and legend are adjusted.
      theme(plot.title = element_text(size = 14, face = "bold", hjust = 0.5),
           plot.subtitle = element_text(hjust = 0.5),
           legend.position = c(0.5,0),
           legend.margin = margin(0,0,-23,0),
           legend.box.margin = margin(0,0,-23,0),
           legend.direction = "horizontal",
           legend.text = element_text(size = 11)) +
           #Filling colors: Online hex identifier is used to obtain hex values.
           scale_fill_manual("",
                     values = c("SoundExchange Distributions"  = "#2d7a7a",
                                "Other Ad-Supported Streaming" = "#c2da74"))
```

Combined our graph with the annual revenues.

```
g +
  #Background and axes are removed.
  #Text of x axis and the title of y axis are resized and positioned properly.
  theme(panel.background = element_blank(),
          axis.ticks = element_blank(),
          axis.line = element_blank(),
          axis.text.y  = element_blank(),
          axis.text.x = element_text(vjust = 5, size = 12),
          axis.title.y = element_text(size = 12)) +
  #X axis  was not at the bottom of the stacked bar,
  #Therefore geom_hline() is used to create an x axis same as the original figure.
  geom_hline(yintercept = 0) +

  #Text for the value of annual revenue is created. Adjusted its text, font and position properly.
  geom_text(aes(Year, Value_Annually,
                label = sprintf("$%d", round(Value_Annually)), fontface=2, fill = NULL),
                size = 4.25, vjust = -0.25, data = radrev_annually)
```

U.S. DIGITAL AND CUSTOMIZED RADIO REVENUES

Source: RIAA