Project 2

CPSC 292 Section #02

11/21/2025

Project 2: Best-Selling Video Games of All Time

Reading Data

```
game.data <- read.csv("../data/best-selling video games of all time.csv")
```

Cleaning Data

```
game.data.clean <- game.data %>%
  clean_names() %>%
  remove_empty(which = c("rows", "cols")) %>%
  na.omit() %>%
  mutate(sales = sales / 100000)

# Adds data frame with median sales for each platform
platforms <- sort(unique(game.data.clean$platform_s))
label_df <- data.frame(platform_s = platforms, median_sales = NA)
```

Cleaning Steps Description:

This cleans the data names, removes any empty columns or rows, and omits any NA data.

Visualizations

Plot 1 with Function and Loop

```
#Plot function
make_plot <- function(data, x_var, y_var, geom_layer) {
  if (is.null(y_var)) {
    ggplot(data, aes_string(x = x_var, fill = x_var)) +
      geom_layer(width = 0.35) +
      theme(
        legend.position = "none",
        axis.title.x = element_text(face = "bold"),
        axis.title.y = element_text(face = "bold"),
```

```
        plot.title = element_text(face = "bold")
      ) +
      coord_flip()
  } else {
    ggplot(data, aes_string(x = x_var, y = y_var, fill = x_var)) +
      geom_layer(width = 0.35) +
      theme(
        legend.position = "none",
        axis.title.x = element_text(face = "bold"),
        axis.title.y = element_text(face = "bold"),
        plot.title = element_text(face = "bold")
      ) +
      coord_flip()
  }
}
#Base plot
plot1 <- make_plot(game.data.clean, "platform_s", "sales", geom_boxplot)
```

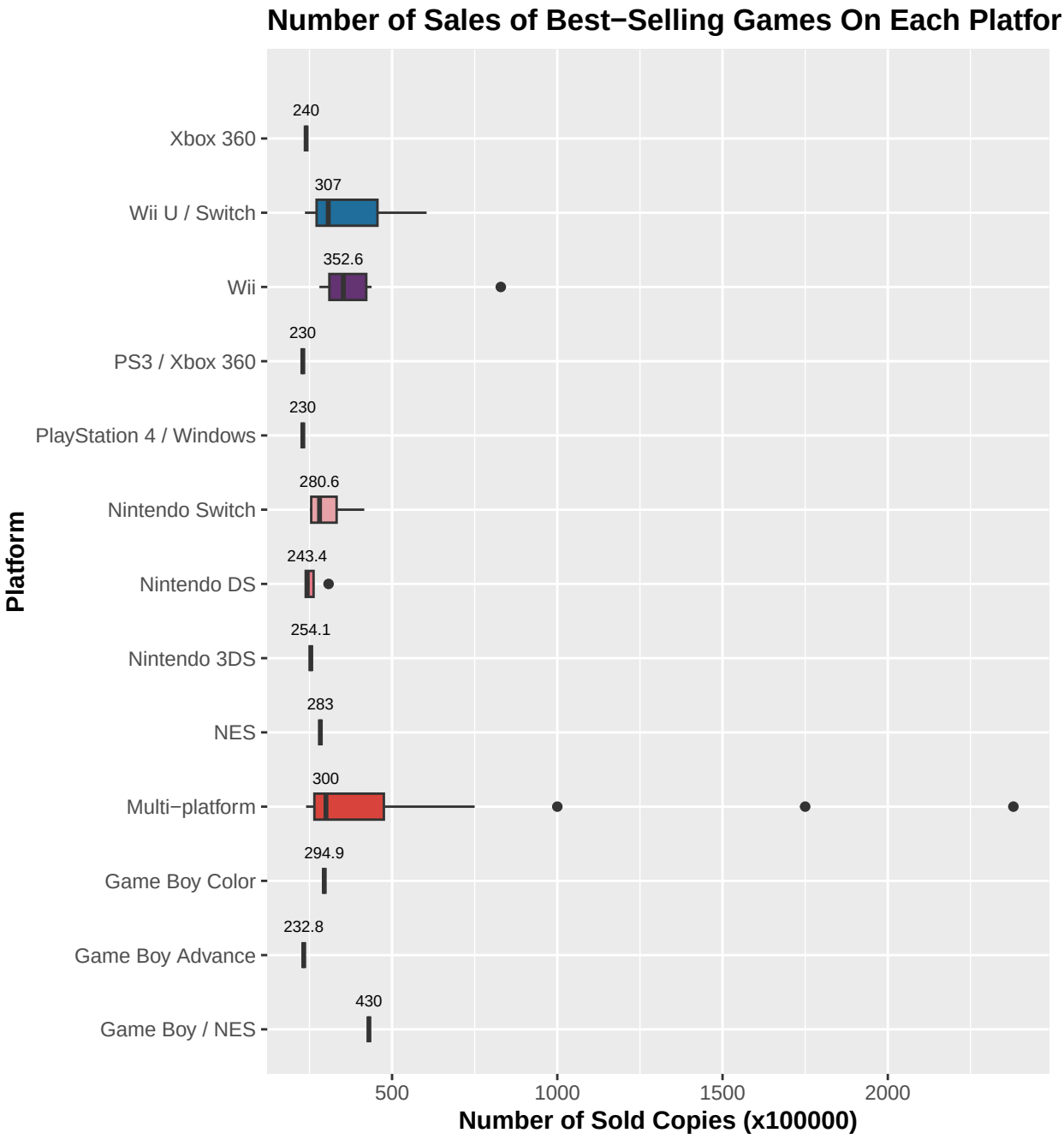
```
## Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation idioms with `aes()`.
## i See also `vignette("ggplot2-in-packages")` for more information.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
plot1 <- plot1 +
  labs(title = "Number of Sales of Best-Selling Games On Each Platform",
      x = "Platform",
      y = "Number of Sold Copies (x100000)") +
  scale_fill_paletteer_d("MetBrewer::Signac") +
  scale_x_discrete(expand = expansion(mult = c(0.05, 0.1)))

#Loop to calculate median sales for each platform
for (i in seq_along(platforms)) {
  p <- platforms[i]
  label_df$median_sales[i] <- median(
    game.data.clean$sales[game.data.clean$platform_s == p],
    na.rm = TRUE
  )
}

label_df
```

```
##                 platform_s median_sales
## 1           Game Boy / NES       430.00
## 2         Game Boy Advance       232.80
## 3           Game Boy Color       294.90
## 4           Multi-platform       300.00
## 5                      NES       283.00
## 6             Nintendo 3DS       254.10
## 7              Nintendo DS       243.45
## 8          Nintendo Switch       280.60
## 9  PlayStation 4 / Windows       230.00
```

```
## 10          PS3 / Xbox 360       230.00
## 11                     Wii       352.60
## 12          Wii U / Switch       307.00
## 13                Xbox 360       240.00
```

```r
#Add median labels
plot1_medians <- plot1 +
  geom_text(
    data = label_df,
    aes(x = platform_s,
        y = median_sales,
        label = round(median_sales, 1)),
    color = "black",
    size = 2.5,
    vjust = -2.0,
    hjust = 0.50
  )

plot1_medians
```




```
ggplot2::ggsave("../results/plot1-function-loop.png", plot = plot1_medians, width = 8, height = 7)
```

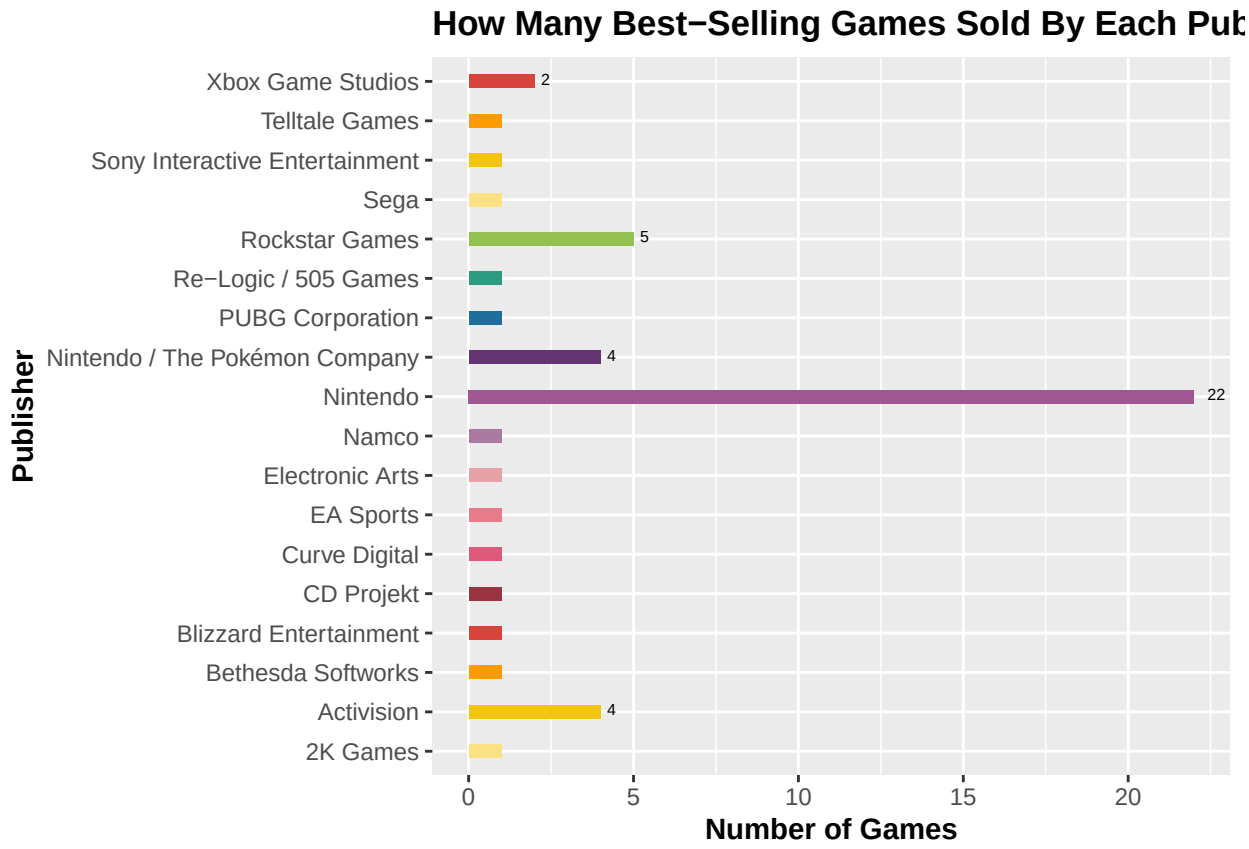
Description of Plot 1 with Function and Loop

This plot shows us different platforms that the most sold video games of all time uses. A "make_plot" function was made to simplify the original code. This function accept arguments including the data set, x variable, y variable, and geom layer style. The function includes a conditional. If the y-variable is provided, it builds a box plot with x and y aesthetics. If the y variable is NULL, it builds a bar plot that only uses the x aesthetic. A loop was used to calculate the median values of number of sold copies (x100000) for each platform, providing a better understanding of which kind of platforms get more sales based on their games.

Median values were added to the plot to add complexity and increase the amount of information that can be communicated through the visualization. This plot produces the same visualization as the original code, however the code was improved with the addition of the function and the loop.

Plot 2 with Function and Conditional

```
#Color palette
color.base <- paletteer::paletteer_d("MetBrewer::Signac")
color <- rep(color.base, length.out = 18)

#Base plot
plot2 <- make_plot(game.data.clean, "publisher_s", NULL, geom_bar) +
  labs(
    title = "How Many Best-Selling Games Sold By Each Publisher",
    x = "Publisher",
    y = "Number of Games",
    fill = "Publisher"
  ) +

#Conditional to add labels on bars where count of numbers of games is > 1
  scale_fill_manual(values = color) +
  geom_text(
    stat = "count",
    aes(label = ifelse(after_stat(count) > 1, after_stat(count), "")),
    vjust = 0.3,
    hjust = -0.75,
    size = 2
  )

plot2
```




```
ggplot2::ggsave("../results/plot2-function-conditional.png", plot = plot2, width = 8, height = 5)
```

Description of Plot 2 with Function and Conditional

This plot shows how many of the best selling video games of all time belong to each publisher. The "make_plot" function was used with the y-variable argument as "NULL". This called the conditional of the function to go through the "else" branch to produce a bar plot.

A conditional was added to the geom text layer to include labels for the number of games sold by each publisher if the value was greater than 1. This feature adds an additional visual element that makes interpreting the data easier. This plot produces the same visualization as the original code, however adding the conditional made the plot more informative by highlighting publishers with many best-selling games.

Responses to Feedback

There were a few changes that were made to the code and data to improve the visualizations. These include renaming the data folder from "game.data" to "data". Additionally, the style was improved and revised according to style guide in Best Practices. Comments were also added to describe each section of code.

References

https://dplyr.tidyverse.org/ https://ianadamsresearch.com/post/using-ggplot2-to-visualize-the-frequency-of-your-name/ https://github.com/thomasp85/patchwork#patchwork https://www.sthda.com/english/

wiki/bar-plots-r-base-graphs https://exts.ggplot2.tidyverse.org/gallery/ https://r-graph-gallery.com/ggplot2-color.html