Final\_Project\_Q2.2\_R\_Script.R

2025-05-03

***## RESEARCH QUESTION 2.2 R-SCRIPT***  
**install.packages("readxl")**

**install.packages("dplyr")**

**install.packages("tidyr")**

**install.packages("tidyverse")**

**install.packages(c("dplyr", "tidyr", "magrittr"))**

**# Load packages**

**library(dplyr)**

**library(tidyr)**

**library(readxl)**

**library(tidyverse)**

**library(magrittr)**

*# Read in the data*  
df <- read\_excel("/Users/arundhathiroy/Downloads/vgsales\_final.xlsx")  
# Define platform types

handhelds <- c("3DS", "DS", "GBA", "PSP", "PSV", "Switch", "GB")

stationary <- c("PS", "PS2", "PS3", "PS4", "PS5", "X360", "XB", "XOne", "NGC", "Wii", "WiiU")

# Define platform types

handhelds <- c("3DS", "DS", "GBA", "PSP", "PSV", "Switch", "GB")

stationary <- c("PS", "PS2", "PS3", "PS4", "PS5", "X360", "XB", "XOne", "NGC", "Wii", "WiiU")

# Add console type classification

df <- df %>%

mutate(Device\_Type = case\_when(

Platform %in% handhelds ~ "Handheld",

Platform %in% stationary ~ "Stationary",

TRUE ~ "Other"

))

# Filter recent decade

df\_recent <- df %>% filter(Year >= 2010 & !is.na(Device\_Type))

# Total global sales by device type

sales\_summary <- df\_recent %>%

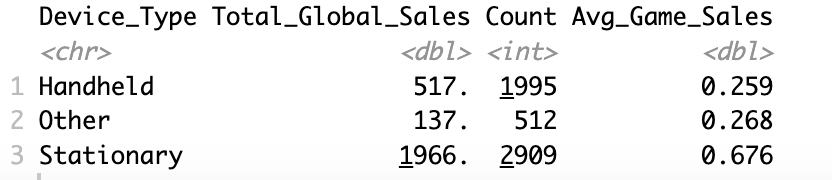
group\_by(Device\_Type) %>%

summarise(Total\_Global\_Sales = sum(Global\_Sales),

Count = n(),

Avg\_Game\_Sales = mean(Global\_Sales))

print(sales\_summary)



# By Region

regional\_summary <- df\_recent %>%

group\_by(Device\_Type) %>%

summarise(

NA\_Sales = sum(NA\_Sales),

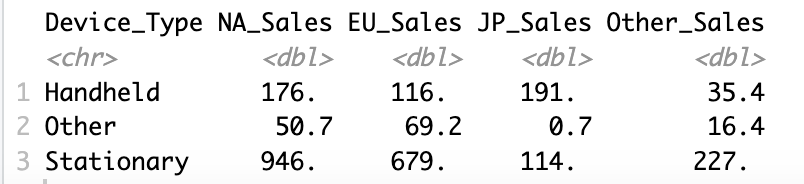
EU\_Sales = sum(EU\_Sales),

JP\_Sales = sum(JP\_Sales),

Other\_Sales = sum(Other\_Sales)

)

print(regional\_summary)



#creating visualizations

# Add Console\_Type and Company columns

df <- df %>%

mutate(

Console\_Type = case\_when(

Platform %in% handheld\_platforms ~ "Handheld",

Platform %in% stationary\_platforms ~ "Stationary",

TRUE ~ "Other"

),

Publisher = ifelse(is.na(Publisher), "Unknown", Publisher),

Company = case\_when(

grepl("Nintendo", Publisher, ignore.case = TRUE) ~ "Nintendo",

grepl("Sony|PlayStation", Publisher, ignore.case = TRUE) ~ "Sony",

grepl("Microsoft|Xbox", Publisher, ignore.case = TRUE) ~ "Microsoft",

TRUE ~ "Other"

)

)

# Summarize sales by company and console type

sales\_summary <- df %>%

group\_by(Company, Console\_Type) %>%

summarise(Global\_Sales = sum(Global\_Sales, na.rm = TRUE), .groups = "drop")

# -----------------------------

# 📊 1. Grouped Bar Chart

# -----------------------------

ggplot(sales\_summary, aes(x = Company, y = Global\_Sales, fill = Console\_Type)) +

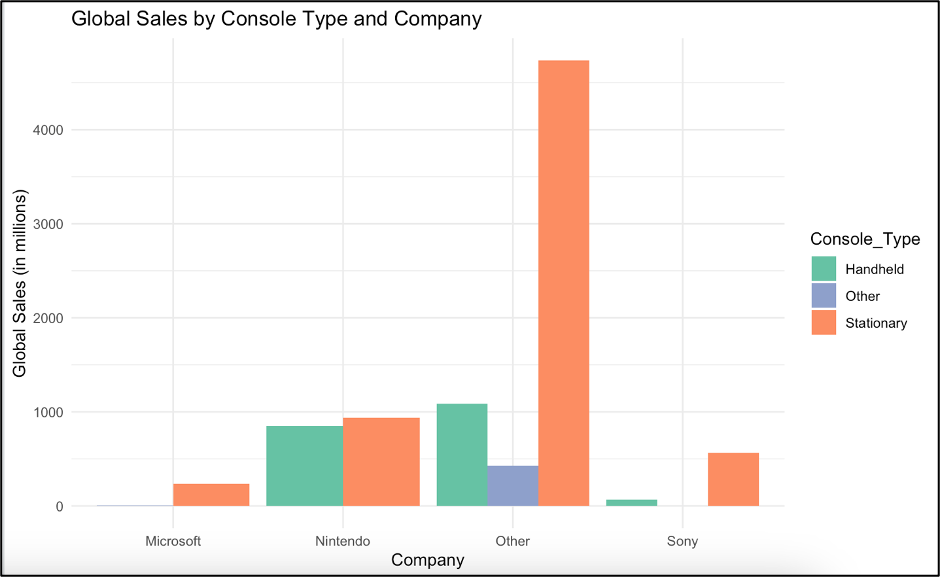
geom\_bar(stat = "identity", position = "dodge") +

labs(title = "Global Sales by Console Type and Company",

x = "Company", y = "Global Sales (in millions)") +

theme\_minimal() +

scale\_fill\_manual(values = c("Handheld" = "#66c2a5", "Stationary" = "#fc8d62", "Other" = "#8da0cb"))



#trend analysis

# Clean and safely convert Year

df <- df %>%

mutate(

Year = as.character(Year), # make sure it's character

Year = ifelse(Year %in% c("N/A", "TBD", ""), NA, Year), # replace non-numeric text with NA

Year = as.numeric(Year) # now safely convert

) %>%

filter(!is.na(Year)) # remove rows where Year couldn't be converted

# Summarize total sales by year and console type

trend\_data <- df %>%

group\_by(Year, Console\_Type) %>%

summarise(Global\_Sales = sum(Global\_Sales, na.rm = TRUE), .groups = "drop")

summary\_df <- df %>%

group\_by(Company, Console\_Type) %>%

summarise(Global\_Sales = sum(Global\_Sales, na.rm = TRUE)) %>%

arrange(desc(Global\_Sales))

#T-test : Stationary Vs Handheld

# Filter only Handheld and Stationary (exclude 'Other')

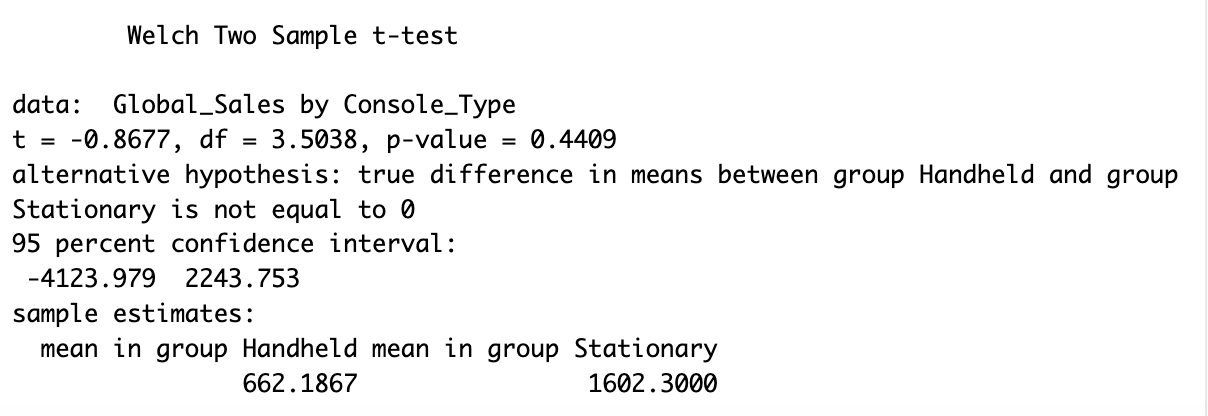
core\_data <- summary\_df %>%

filter(Console\_Type %in% c("Handheld", "Stationary"))

# Run independent t-test

t\_test\_result <- t.test(Global\_Sales ~ Console\_Type, data = core\_data)

print(t\_test\_result)



anova\_model <- aov(Global\_Sales ~ Console\_Type, data = summary\_df)

summary(anova\_model)

