

Assignment 5

Sophie Giacobbe

2/23/2023

```
#load in necessary packages
```

```
library(dplyr)
library(ggplot2)
library(ggrepel)
library(readr)
```

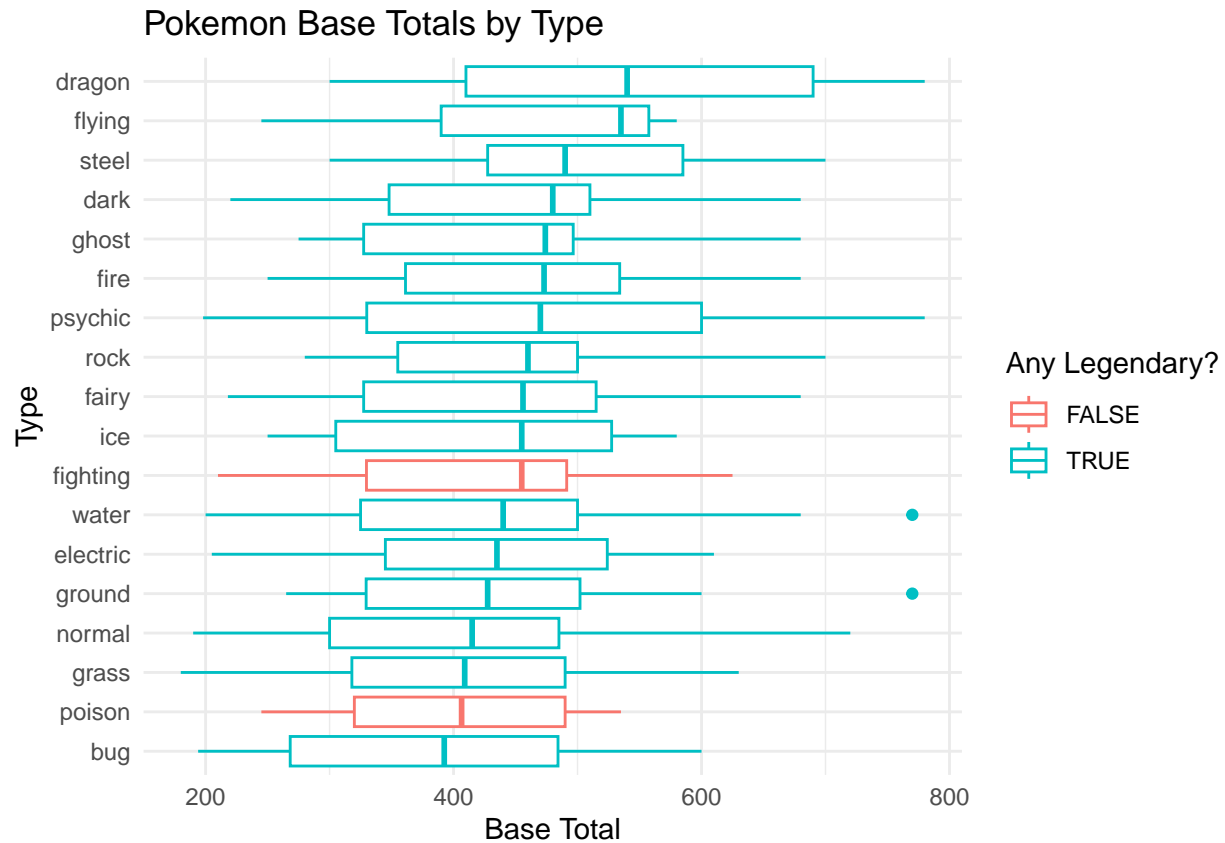
```
#read in Pokemon data
```

```
pokemon <- read.csv("pokemon.csv")
firstgen <- pokemon|>
  filter(generation == 1) |>
  select(-(abilities:against_water), -(base_egg_steps:base_happiness)) |>
  mutate(type1 = ifelse(type1 == "fairy", "normal", type1))
```

```
#show distribution of base stats by type
```

```
pokemon <- pokemon |>
  group_by(type1) |>
  mutate(anyleg = any(isLegendary))

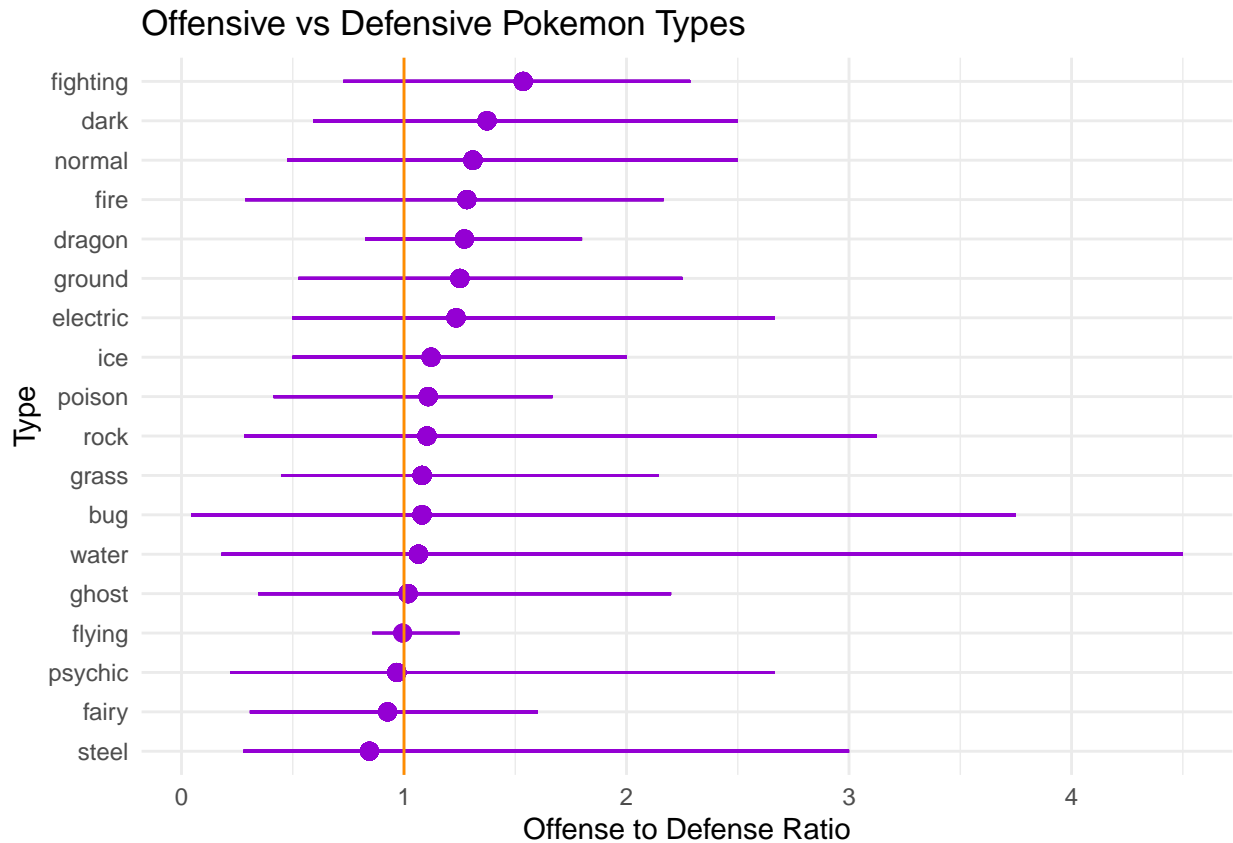
ggplot(pokemon) +
  geom_boxplot(aes(x = reorder(type1, base_total, median, na.rm = TRUE), y = base_total,
    color = anyleg)) +
  coord_flip() +
  labs(x = "Type",
    y = "Base Total",
    title = "Pokemon Base Totals by Type") +
  guides(color = guide_legend(title = "Any Legendary?"))+
  theme_minimal()
```



```
#which types skew offensive, which skew defensive?
pokemon <- pokemon |>
  group_by(type1) |>
  mutate(ratio = attack / defense)

pokemon2 <- pokemon |>
  group_by(type1) |>
  summarize(ratio) |>
  mutate(pmax = max(ratio)) |>
  mutate(pmin = min(ratio)) |>
  mutate(pmean = mean(ratio))

ggplot(pokemon2) +
  geom_pointrange(aes(y = reorder(type1, ratio), x = pmean, xmin = pmin, xmax = pmax),
    color = "darkviolet") +
  geom_vline(aes(xintercept = 1), color = "darkorange") +
  theme_minimal() +
  labs(x = "Offense to Defense Ratio",
    y = "Type",
    title = "Offensive vs Defensive Pokemon Types")
```



```
#read in video game sales data
vgsales <- read_csv("vgsales.csv")

gameboy_sales <- vgsales |>
  filter(Platform == "GB") |>
  filter(NA_Sales > 0, JP_Sales > 0) |>
  mutate(NA_Sales = NA_Sales * 1000000,
         JP_Sales = JP_Sales * 1000000,
         Global_Sales = Global_Sales * 1000000)

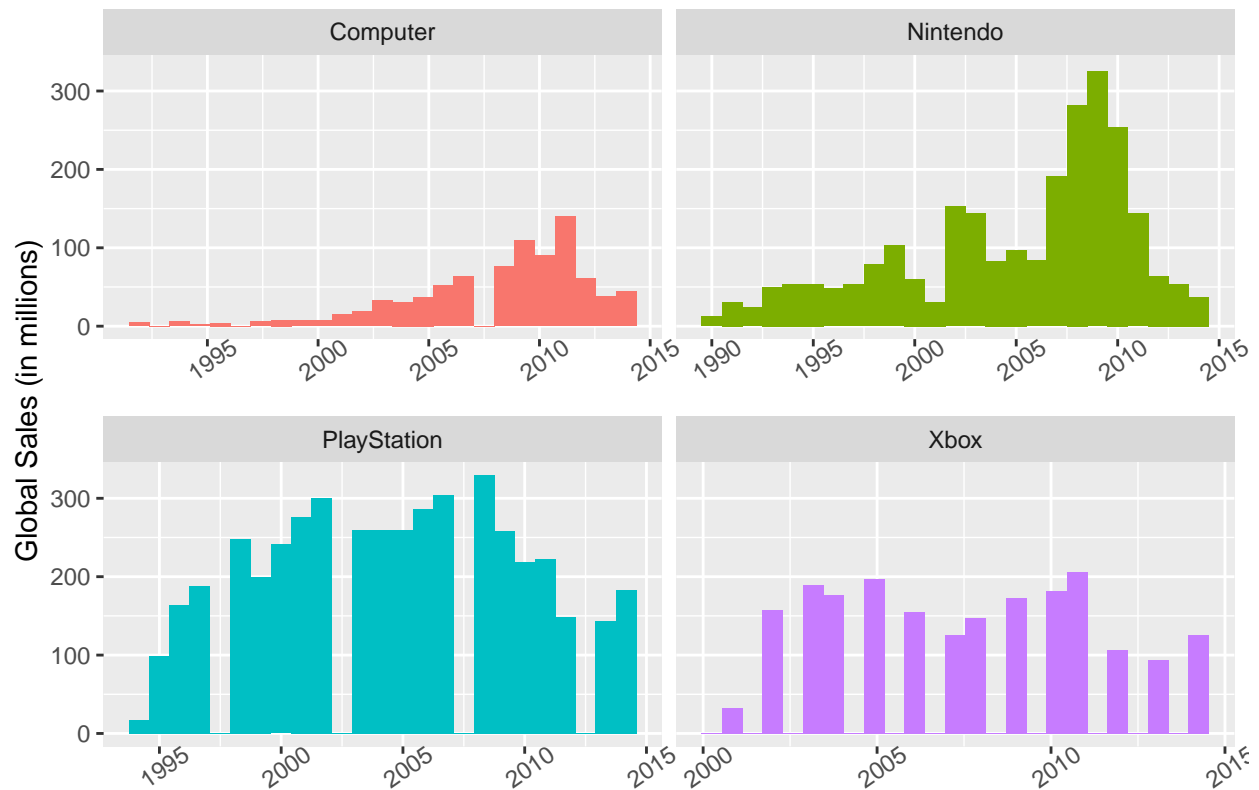
playstation <- c("PS", "PS2", "PS3", "PS4")
xbox <- c("XB", "X360", "XOne")
nintendo <- c("NES", "SNES", "N64", "GC", "Wii", "WiiU")

top_platform_sales <- vgsales |>
  mutate(platform2 = case_when(Platform %in% playstation ~ "PlayStation",
                              Platform %in% xbox ~ "Xbox",
                              Platform %in% nintendo ~ "Nintendo",
                              Platform == "PC" ~ "Computer",
                              TRUE ~ "Other"),
         Year = as.numeric(Year)) |>
  filter(platform2 != "Other", !is.na(Year), Year <= 2014, Year >= 1990)

#understand how sales for the top 4 video game platforms change across time
top_platform_sales <- top_platform_sales |>
  mutate(Sales = NA_Sales + JP_Sales + EU_Sales + Other_Sales + Global_Sales) |>
```

```
group_by(platform2, Year)

ggplot(top_platform_sales) +
  geom_histogram(aes(x = Year, fill = platform2), bins = 25) +
  facet_wrap("platform2", scales = "free_x") +
  theme(axis.text.x = element_text(angle = 35), legend.position = "none") +
  labs(x = "", y = "Global Sales (in millions)")
```



```
#display differences in North American and Japanese gameboy sales
ggplot(gameboy_sales, aes(NA_Sales, JP_Sales)) +
  geom_point(aes(color = "orange")) +
  geom_label_repel(data = gameboy_sales |>
    filter(Global_Sales > 10000000), mapping = aes(label = Name, size = 3),
    color = "maroon", min.segment.length = 0) +
  scale_y_log10(labels = scales::comma) +
  scale_x_log10(labels = scales::comma) +
  labs(x = "North American Sales", y = "Japanese Sales", title = "Gameboy Sales") +
  theme(legend.position = "none")
```

Gameboy Sales

