

Information Science 3

Final Project

Nobuoka Yuki (1240492)

2022-11-30

Introduction

The story I came up with is “Get more medals at the Olympics held in home country than when they are held in other countries”. I believe that behind this is the low physical burden of traveling, experiential knowledge of the climate and topography, and there is a sports enhancement strategy that has been in place since decided to host. I will examine the story conceived from the above background using three types of diagrams.

Data

Data were taken from OLYMPICS_athlete_events.csv created Randi H. Griffin. The data are individual level performances from the Winter and Summer Olympics from 1896 to 2016. This time, I used data from 1980.

```
#データの読み込み
data <- read.csv("OLYMPICS_athlete_events.csv")
#1980 年以降の大会を抽出
data <- data %>%
  filter(Year >= 1980) %>%
  select(Year, NOC, City, Medal, Sex, Season) %>%
  arrange(Year)
#開催国と選手の出身国の表記を統一
data <- data %>%
  mutate(City = str_replace_all(City, pattern = c("Lake Placid" = "USA",
```

```

"Moskva" = "RUS",
"Sarajevo" = "YUG",
"Los Angeles" = "USA",
"Calgary" = "CAN",
"Seoul" = "KOR",
"Barcelona" = "ESP",
"Albertville" = "FRA",
"Lillehammer" = "NOR",
"Atlanta" = "USA",
"Nagano" = "JPN",
"Sydney" = "AUS",
"Salt Lake City" = "USA",
"Athina" = "GRE",
"Torino" = "ITA",
"Beijing" = "CHN",
"Vancouver" = "CAN",
"London" = "GBR",
"Sochi" = "RUS",
"Rio de Janeiro" = "BRA"))))

```

#開催国か否かの変数を追加

```

data <- data %>%
  mutate(Situation = ifelse(NOC == City, "Host", "Guest"))

```

#冬と夏を区別（文字列を操作）

```

data <- data %>%
  # mutate(Year_place = paste(!!!rlang::syms(c("Year", "Season")), sep = "_"))
  mutate(Year_place = str_c(Year, Season, sep = "_"))

```

Visualization

Box plot

Visualize the number of medals for each country that has experienced hosting the Olympics in its own country since 1980 with a box plot.

```

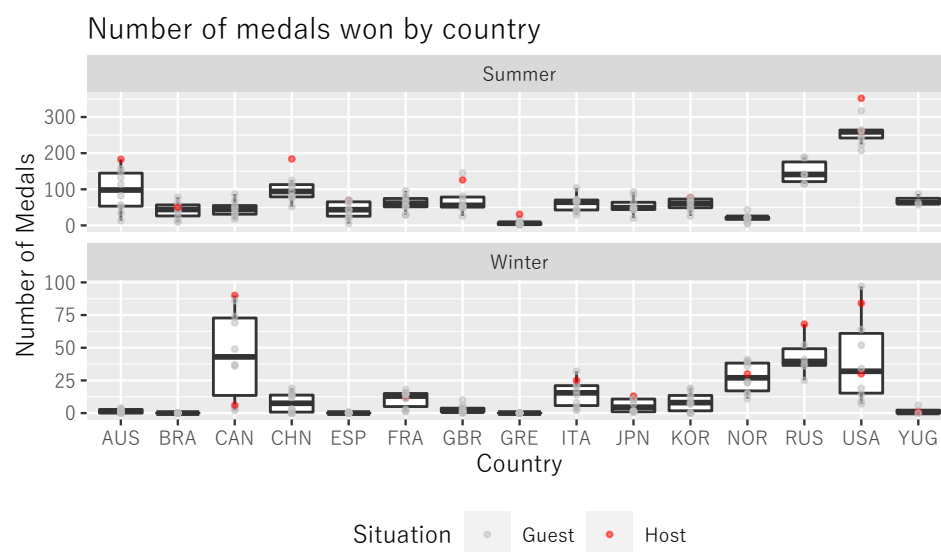
Year_Medal_data <- data %>%
  group_by(Year, Year_place, NOC) %>%
  # 年と国ごとにメダル数をカウント
  summarise(Medal_count = sum(ifelse(is.na(Medal), 0, 1)), #NAがあるときは is.na()
            Situation = Situation[1], #これの方が分かりやすい
            Sensyu = length(NOC),
            Medal_per = (Medal_count / Sensyu) * 100, #1人当たりのメダル獲得数
            Season = max(Season),
            .groups = "drop")
Year_Medal_data <- Year_Medal_data %>% #開催したことのある国だけを抽出
  filter(NOC %in% unique(data$City))

```

```

Year_Medal_data %>%
  ggplot() +
  geom_boxplot(aes(x = NOC, y = Medal_count),
               outlier.shape = NA, size = 0.5) +
  geom_point(aes(x = NOC, y = Medal_count, color = Situation),
             alpha = 0.5, size = 0.7) +
  scale_colour_manual(values = c("gray", "red")) +
  theme(legend.position = "bottom") +
  facet_wrap(~ Season, ncol = 1, scale = "free_y") +
  labs(x = "Country", y = "Number of Medals") +
  ggtitle("Number of medals won by country")

```



The horizontal axis is the countries that have experienced the Olympics, and the vertical axis is the number of medals won in each competition. The black dots represent competition held in other countries, and the red dots represent competition held in the home country. It can be seen that the number of medals won is actually higher when host countries.

Bar graph

Create a bar graph to compare the average number of medals got for each competition, whether held in another country or held in your own country.

#データの入れ物を用意

```
Year_Medal_data4 <- tibble(  
  Season = rep("Season", 4*length(unique(Year_Medal_data$NOC))),  
  NOC = rep("NOC", 4*length(unique(Year_Medal_data$NOC))),  
  Situation = rep("Kaisai", 4*length(unique(Year_Medal_data$NOC))),  
  Medal_count = rep(0, 4*length(unique(Year_Medal_data$NOC)))  
)
```

```
n <- 1
```

```
for (i in unique(Year_Medal_data$NOC)) {  
  df1 <- Year_Medal_data %>%  
    filter(NOC == i)  
  
  #夏開催で非開催地  
  df2 <- df1 %>%  
    filter(Season == "Summer")  
  df3 <- df2 %>%  
    filter(Situation == "Guest")  
  S_G_medal <- round(mean(df3$Medal_count), digits = 2)  
  Year_Medal_data4[n, 1] <- "Summer"  
  Year_Medal_data4[n, 2] <- i  
  Year_Medal_data4[n, 3] <- "Guest"  
  Year_Medal_data4[n, 4] <- S_G_medal
```

#夏開催で開催地

```
df3 <- df2 %>%
```

```

    filter(Situation == "Host")
  S_H_medal <- round(mean(df3$Medal_count), digits = 2)
  Year_Medal_data4[n + 1, 1] <- "Summer"
  Year_Medal_data4[n + 1, 2] <- i
  Year_Medal_data4[n + 1, 3] <- "Host"
  Year_Medal_data4[n + 1, 4] <- S_H_medal

  #冬開催で非開催地
  df2 <- df1 %>%
    filter(Season == "Winter")
  df3 <- df2 %>%
    filter(Situation == "Guest")
  W_G_medal <- round(mean(df3$Medal_count), digits = 2)
  Year_Medal_data4[n + 2, 1] <- "Winter"
  Year_Medal_data4[n + 2, 2] <- i
  Year_Medal_data4[n + 2, 3] <- "Guest"
  Year_Medal_data4[n + 2, 4] <- W_G_medal

  #冬開催で開催地
  df3 <- df2 %>%
    filter(Situation == "Host")
  W_H_medal <- round(mean(df3$Medal_count), digits = 2)
  Year_Medal_data4[n + 3, 1] <- "Winter"
  Year_Medal_data4[n + 3, 2] <- i
  Year_Medal_data4[n + 3, 3] <- "Host"
  Year_Medal_data4[n + 3, 4] <- W_H_medal

  n <- n + 4
}

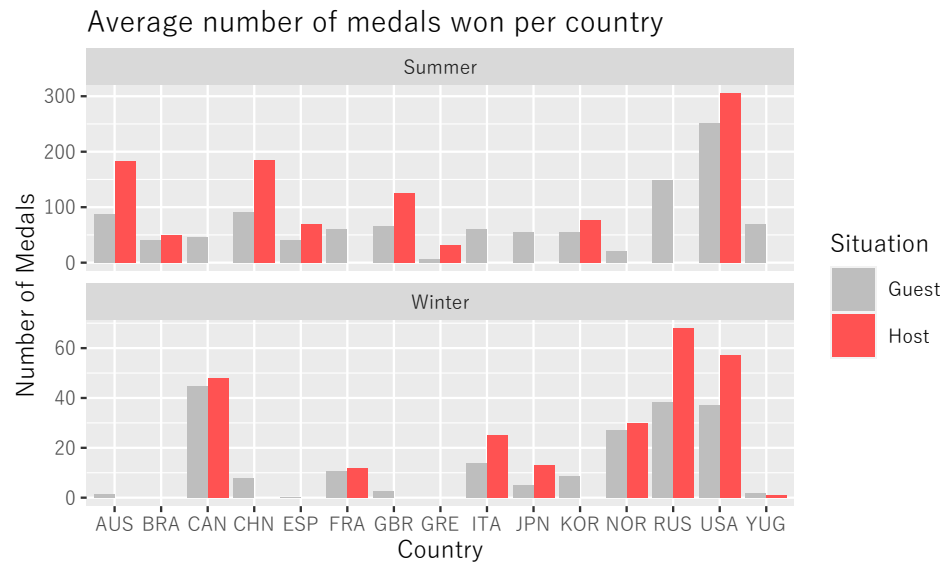
```

```

Year_Medal_data4 %>%
  ggplot() +
  geom_bar(aes(x = NOC, y = Medal_count, fill = Situation),
           stat = "identity", position = "dodge") +
  scale_fill_manual(values = c("gray", "#ff5252")) +
  facet_wrap(~ Season, ncol = 1, scales = "free_y") +
  labs(x = "Country", y = "Number of Medals") +

```

```
ggtitle("Average number of medals won per country")
```

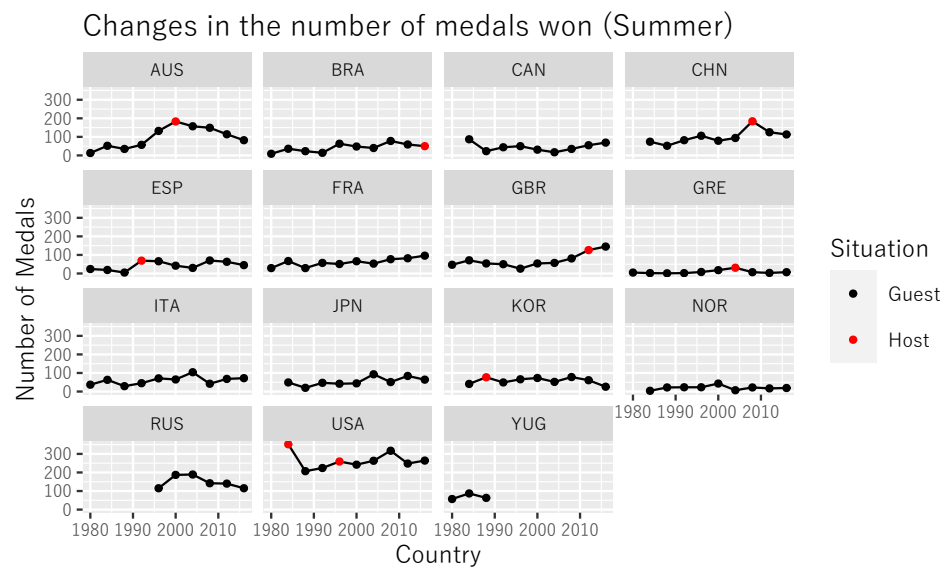


The horizontal axis shows the countries that have hosted the Olympics, and the vertical axis shows the average number of medals won per competition. Red indicates when the event is held in another country, and blue indicates when it is held in your own country. From this figure, it can be seen that the average number of medals won by almost all host countries is higher than when other countries hosted.

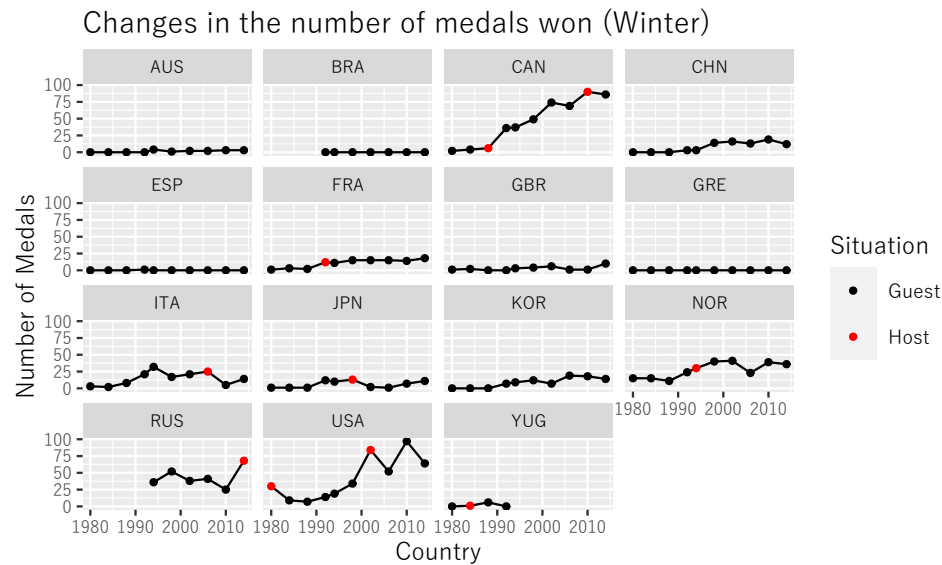
Line chart

Use a line chart to check how the number of medals acquired changes over time.

```
Year_Medal_data %>%
  filter(Season == "Summer") %>%
  ggplot() +
  geom_line(aes(x = Year, y = Medal_count, group = 1), size = 0.4) +
  geom_point(aes(x = Year, y = Medal_count, color = Situation), size = 0.8) +
  scale_colour_manual(values = c("black", "red")) +
  facet_wrap(~ NOC, ncol = 4) +
  labs(x = "Country", y = "Number of Medals") +
  ggtitle("Changes in the number of medals won (Summer)") +
  theme(strip.text = element_text(size = rel(0.7), color = "black"),
        axis.text = element_text(size = rel(0.7)))
```



```
Year_Medal_data %>%
  filter(Season == "Winter") %>%
  ggplot() +
  geom_line(aes(x = Year, y = Medal_count, group = 1), size = 0.4) +
  geom_point(aes(x = Year, y = Medal_count, color = Situation), size = 0.8) +
  scale_colour_manual(values = c("black", "red")) +
  facet_wrap(~ NOC, ncol = 4) +
  labs(x = "Country", y = "Number of Medals") +
  ggtitle("Changes in the number of medals won (Winter)") +
  theme(strip.text = element_text(size = rel(0.7), color = "black"),
        axis.text = element_text(size = rel(0.7)))
```



The horizontal axis is the year of the Olympics and the vertical axis is the number of medals won. The black dots indicate that the Olympics were held in another country, and the red dots indicate that the Olympics were held in the home country.

From this figure, it is possible to read two patterns of chronological changes in the number of medals won. The first is a pattern in which the number of medals won has been increasing since before the home country hosted the competition (e.g. Winter CAN, Winter USA, Summer GBA), and the second is a pattern in which the number of medals won has been decreasing but rose significantly in the year the country hosted the competition (e.g. Summer RUS, Summer CHN). In the former pattern, it is thought that the decision to host the games in the home country led to the implementation of a strategy to strengthen sports at the national level, which led to an increase in the number of medals won. In the latter pattern, the number of medals won was on a downward trend before the home country hosted the event, but when the home country hosted the event, it was superior in terms of reduced travel burden and empirical knowledge of the climate and topography, so the number of medals won dramatically.

Conclusion

These three graphs reflect the story I thought. In addition, the line graph in particular can read sports enhancement strategy that has been in place since decided to host, which is the background of the story. My story is also consistent with the visualized data, suggesting correct possibilities.