Final Project

Yifei Yang, Yiming Yuan

Read in the data

```
library(readr)
ufc <- read_csv("data/ufc-fighters-statistics.csv")
```

Introduction and data

Around 300 million individuals worldwide identify themselves as fans of Mixed Martial Arts (MMA), with its popularity peaking in nations such as the United States, the United Kingdom, Brazil, Singapore, and China. The Ultimate Fighting Championship (UFC) is the premier organization in the MMA world. Our motivation is to research on what contributes to the fighters' performance. Thus, the research questions are: How largely do the innate physical attributes affect fighters' performance? What's the most effective stance? How does the way fighters stroke (speed/accuracy/amount/defence success) influence their performance?

Today's data are UFC fighter statistics including:

wins: The number of wins the fighter has in their career.

draws: The number of draws the fighter has in their career.

height_cm: The height of the fighter in centimeters.

weight_in_kg: The weight of the fighter in kilograms.

reach_in_cm: The reach of the fighter in centimeters.

stance: The fighting stance of the fighter (Orthodox/Southpaw/Switch).

significant_strikes_landed_per_minute: The average number of significant strikes landed by the fighter per minute.

significant_striking_accuracy: The percentage of significant strikes that land successfully for the fighter.

significant_strikes_absorbed_per_minute: The average number of significant strikes absorbed by the fighter per minute.

significant_strike_defence: The percentage of opponent's significant strikes that the fighter successfully defends.

average_takedowns_landed_per_15_minutes: The average number of takedowns landed by the fighter per 15 minutes.

takedown_accuracy: The percentage of takedown attempts that are successful for the fighter.

takedown_defense: The percentage of opponent's takedown attempts that the fighter successfully defends.

average_submissions_attempted_per_15_minutes: The average number of submission attempts made by the fighter per 15 minutes.

Our response variable is wins, predictors are

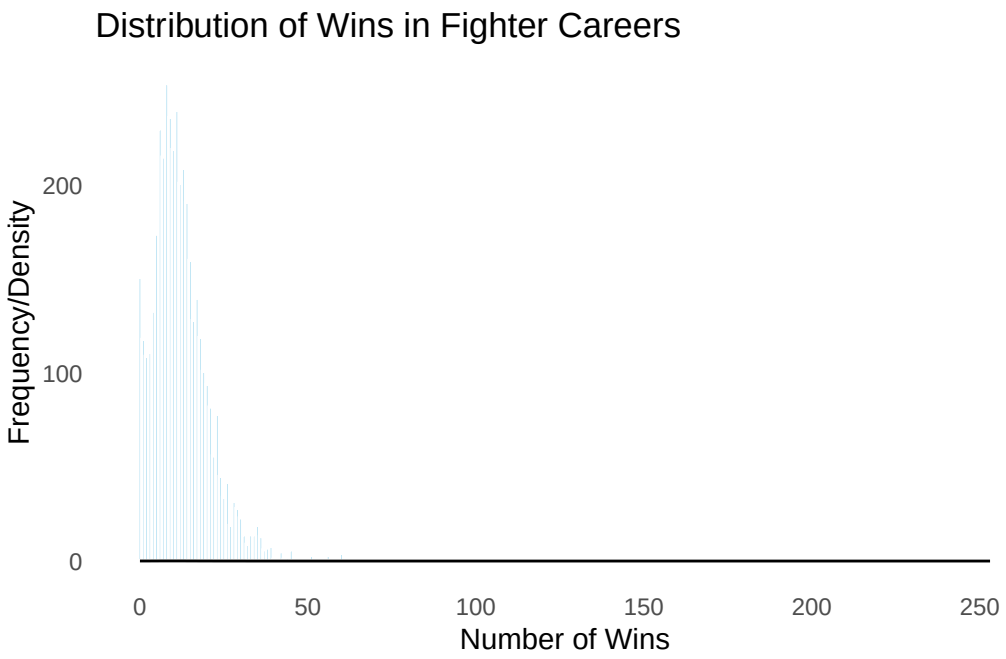
sources:

https://www.kaggle.com/datasets/aaronfriasr/ufc-fighters-statistics?resource=download

https://www.euronews.com/business/2023/09/27/the-booming-billion-dollar-business-of-combat-sports

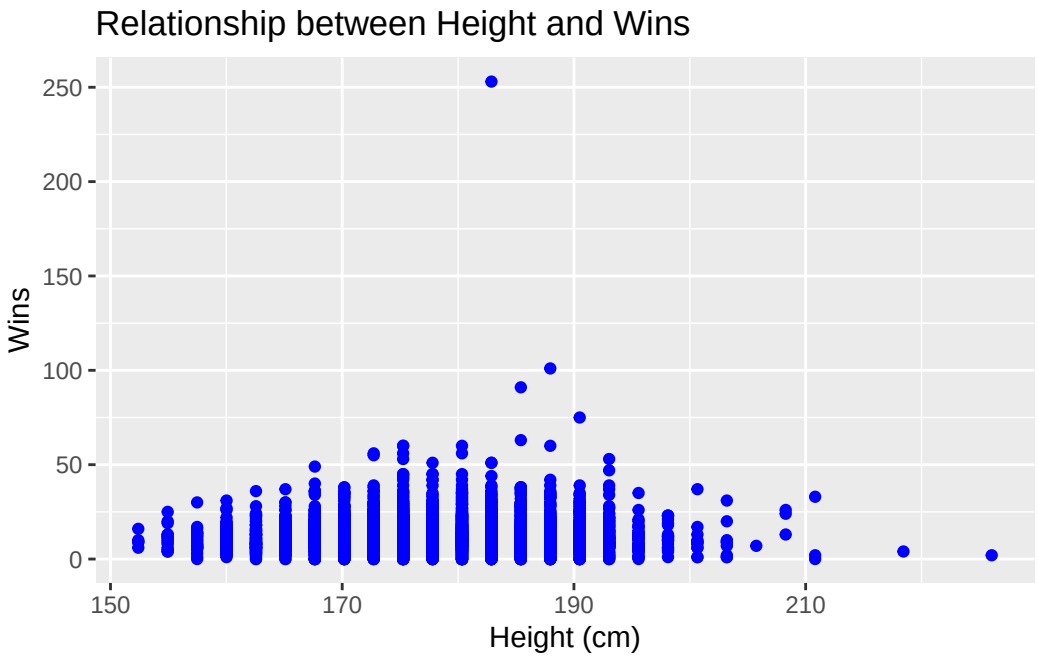
EDA

```
library(ggplot2)

plot <- ggplot(ufc, aes(x = wins)) +
        geom_histogram(binwidth = 1, fill = "skyblue", color = "white") +
        geom_density(alpha = 0.5, fill = "skyblue") +
        labs(x = "Number of Wins", y = "Frequency/Density",
             title = "Distribution of Wins in Fighter Careers") +
        theme_minimal() +
        theme(panel.grid = element_blank())

print(plot)
```



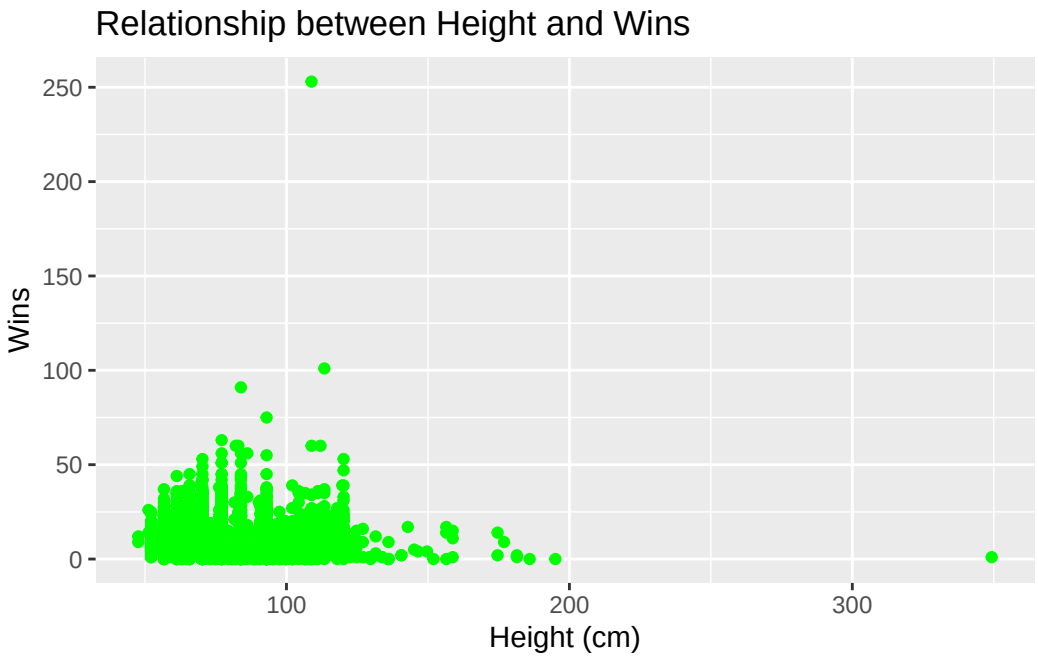

```
plot_height <- ggplot(ufc, aes(x = height_cm, y = wins)) +
               geom_point(color = "blue") +
               labs(x = "Height (cm)", y = "Wins",
                    title = "Relationship between Height and Wins")

print(plot_height)
```

```
Warning: Removed 298 rows containing missing values (`geom_point()`).
```



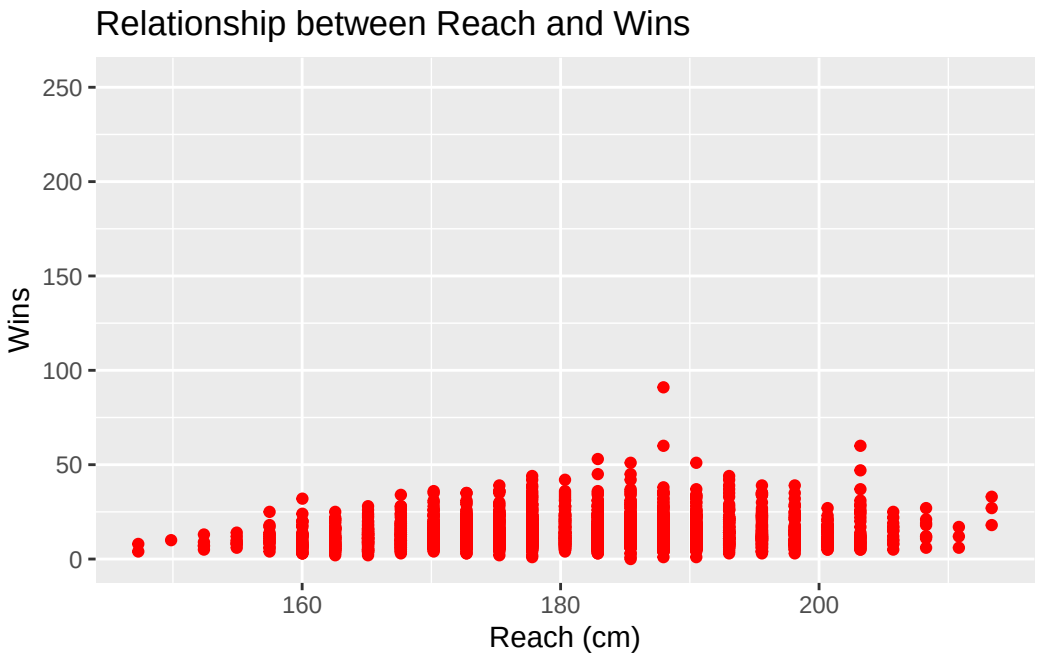

```
plot_weight <- ggplot(ufc, aes(x = weight_in_kg, y = wins)) +
               geom_point(color = "green") +
               labs(x = "Height (cm)", y = "Wins",
                    title = "Relationship between Height and Wins")

print(plot_weight)
```

```
Warning: Removed 87 rows containing missing values (`geom_point()`).
```




```
plot_reach <- ggplot(ufc, aes(x = reach_in_cm, y = wins)) +
              geom_point(color = "red") +
              labs(x = "Reach (cm)", y = "Wins",
                   title = "Relationship between Reach and Wins")
print(plot_reach)
```

```
Warning: Removed 1927 rows containing missing values (`geom_point()`).
```


Relationship between Reach and Wins
Wins
250
200
150
100
50
0
160
180
200
Reach (cm)