Impact of Age and Environmental Conditions on Marathon Performance

Abstract

This study investigates the influence of age, sex, and environmental factors on marathon performance, utilizing data from five major marathons over a period of 15-20 years. We aimed to address three specific objectives: AIM 1 examined the effects of increasing age on performance differences between men and women, identifying significant variations in how aging impacts athletes differently based on gender. AIM 2 explored the impact of environmental conditions such as temperature and humidity, encapsulated by the Wet Bulb Globe Temperature (WBGT) and other factors, on performance across different age groups and sexes, highlighting how these influences vary among demographic groups. AIM 3 focused on identifying which specific weather parameters, including WBGT, temperature, and flag conditions, most significantly affect marathon outcomes. Through regression analysis and machine learning, our study provides crucial insights into how these variables interplay to affect marathon performance, which is vital for developing strategies to enhance athlete performance under adverse environmental conditions. This comprehensive approach helps in crafting targeted training and safety measures, particularly under challenging weather conditions.

1. Introduction

Marathon has many physiological challenges influenced by both internal and external factors. Based on previous study by Dr. Brett Romano Ely and Dr. Matthew Ely, this project delves into relationship ages, environment, and marathon performance. The demands of marathons provide context for studying human endurance. Marathon performance is determined by an athlete's training but also influenced by their ability to adapt various environmental conditions. Research has shown that endurance is negatively affected by hot conditions, which elevate body temperature and stress systems. Such effects are increased over longer distances due to decreased resilience.

Additionally, research also shows that sex differences in thermoregulatory responses. Studies suggest that women generally have a higher surface area-to-mass ratio, which can affect heat dissipation but also means a quicker onset of heat storage. Moreover, hormonal variations across sex and age play critical roles in determining vascular responses and sweat rates during endurance activities. Understanding these differences is crucial for developing training programs that enhance performance and safety. The goal of this study is to evaluate how different age groups and sexes perform in marathon under different environmental conditions. This involves analysis of historical data spanning 15-20 years from five major marathons. By examining top performances, this study seeks to find the clear patterns on how environmental factors interact with biological factors.

This study fills the gaps in our understanding of endurance, particularly in how demographic variables modulate the effects of environmental stressors on performance. Insights from this research are expected to inform not only athletic training regimes but also guidelines for marathon organizations, such as optimal scheduling and support services to mitigate the adverse effects of harsh weather conditions. The findings from this research could have broader implications for public health, particularly in the context of global climate change. As temperatures worldwide continue to rise, understanding how heat affects physical performance and safety becomes increasingly important. This project will contribute valuable knowledge that can help predict and mitigate the health impacts of environmental stressors on vulnerable populations, including older adults and female athletes, who may face greater risks under extreme conditions.

In summary, we seek to use empirical data and analytic techniques to advance our understanding of the relationship between environment, demography, and physical performance. We will get better understanding to promoting performance in endurance sports and changing global climates.

2. Data Description

```
 Race..0.Boston..1.Chicago..2.NYC..3.TC..4.D.      Year      Sex..0.F..1.M.       Flag
 Min.   :0.000                                 Min.   :1993   Min.   :0.0000   Length:11073
 1st Qu.:1.000                                 1st Qu.:2002   1st Qu.:0.0000   Class :character
 Median :2.000                                 Median :2006   Median :1.0000   Mode  :character
 Mean   :1.913                                 Mean   :2006   Mean   :0.5288
 3rd Qu.:3.000                                 3rd Qu.:2012   3rd Qu.:1.0000
 Max.   :4.000                                 Max.   :2016   Max.   :1.0000
    Age..yr.          X.CR              Td..C            Tw..C              X.rh              Tg..C
 Min.   :14.00   Min.   : -2.251   Min.   : 2.00   Min.   :-1.267   Min.   : 0.2781   Min.   : 9.507
 1st Qu.:31.00   1st Qu.: 18.987   1st Qu.: 8.55   1st Qu.: 5.376   1st Qu.: 0.6914   1st Qu.:19.475
 Median :46.00   Median : 36.015   Median :12.50   Median : 8.457   Median :53.0000   Median :24.956
 Mean   :46.47   Mean   : 48.783   Mean   :13.32   Mean   : 9.420   Mean   :42.3246   Mean   :24.945
 3rd Qu.:61.00   3rd Qu.: 63.203   3rd Qu.:17.25   3rd Qu.:13.733   3rd Qu.:64.0000   3rd Qu.:30.000
 Max.   :91.00   Max.   :368.507   Max.   :28.14   Max.   :21.600   Max.   :98.3333   Max.   :44.450
    SR.W.m2            DP               Wind              WBGT              Sex
 Min.   :141.4   Min.   :-7.433   Min.   : 0.000   Min.   : 1.347   Female:5218
 1st Qu.:368.4   1st Qu.: 0.375   1st Qu.: 7.333   1st Qu.: 8.688   Male  :5855
 Median :512.7   Median : 5.086   Median :10.000   Median :12.676
 Mean   :513.8   Mean   : 5.461   Mean   : 9.945   Mean   :12.914
 3rd Qu.:622.6   3rd Qu.:11.000   3rd Qu.:12.200   3rd Qu.:17.740
 Max.   :909.5   Max.   :20.333   Max.   :21.750   Max.   :25.130
```

The dataset is collected from top single-age performances at five major marathons over 15 to 20 years. This collection includes participants ranging from 14 to 85 years of age, offering a framework to see the impact of age and sex alongside environmental conditions on marathon performance. Each entry in the dataset has the performance of an individual participant at a specific marathon, cataloged by the year and location. The structure includes several key variables: race identification categorized by major marathon cities (Boston, Chicago, NYC, TC, and D), which helps in analyzing location wise environmental impacts and trends; the year of the marathon, which is crucial in tracking temporal trends; the participant's sex, providing insight into sex-specific physiological responses and performance; the participant's age, which is crucial for exploring age-related performance metrics and trends; and the core performance metric (%CR), which offers a standardized measure of individual performance relative to the best recorded time.

Performance metrics (%CR) vary significantly, with the best performances showing high variability and ranging from -2.251 to 368.507, suggesting some outliers or extreme values possibly related to exceptional or unusually poor performances. Environmental factors such as temperature, dew point (DP), relative humidity (X.rh), and Wet Bulb Globe Temperature (WBGT) are thoroughly recorded, offering insights into how different weather conditions impact marathon outcomes. For example, WBGT ranges from 1.347 to 25.130, indicating marathons took place under a wide spectrum of thermal stresses.

The dataset also has a range of environmental conditions during the marathons, such as temperature, which illustrates the heat stress participants faced. And also, the relative humidity, enriching the temperature data by offering insights into air moisture content which significantly impacts thermoregulatory responses. Furthermore, globe temperature, reflecting the effect of air temperature. Solar radiation, which quantifies the level of solar energy impacting the runners, and additional metrics like dew point and wind speed, which provide further context on the cooling potential and evaporative cooling capacity during the races. The granularity of the dataset supports a multifaceted analysis approach. Performance trends are examined across different ages and sexes under varied environmental conditions, employing robust methods to draw correlations and develop predictive models.

The analytical approach leverages tools such as R for data and statistical analysis. We also use machine learning algo to get feature importance based on the comprehensive dataset. In conclusion, this dataset not only serves as a foundation for examination of the factors but also provides valuable insights to broader contexts in sports science, event management, and public health policy.

3. Methodology

In our methodology, we use R to perform EDA, which allowed us to delve into the dataset and show relationships, trends, and anomalies. The use of scatter plots in our analysis was instrumental in assessing the relationships of individual variables with marathon performance. This step was crucial for spotting trends and outliers that could potentially influence the subsequent stages of our analysis.

Furthermore, we also employed regression models for a more detailed multivariate analysis. This approach enabled us to explore the complex interactions among multiple variables, such as age, sex, and various environmental conditions, and their collective impact on marathon performance. By applying both linear and non-linear regression techniques, we could model these relationships accurately, providing a robust statistical foundation to support our hypotheses about the causal effects and predictors of performance in marathons. To further enhance our analysis, we use machine learning techniques to evaluate feature importance. This involved using random forests and boosting methods, which are particularly effective for handling the complexities of our dataset, including non-linear interactions and high dimensionality. These machine learning models allowed us to identify the most significant predictors of marathon performance, offering deeper insights into how various demographic and environmental factors contribute to the outcomes of marathon races.

Overall, our methodological approach combined traditional statistical techniques with advanced machine learning models, enabling a comprehensive analysis of the factors that influence marathon performance. This combination of methods not only provided a thorough understanding of the dataset but also ensured that our findings were robust and well-supported by empirical evidence.

4. Results

AIM 1: Age and Sex Effects on Performance

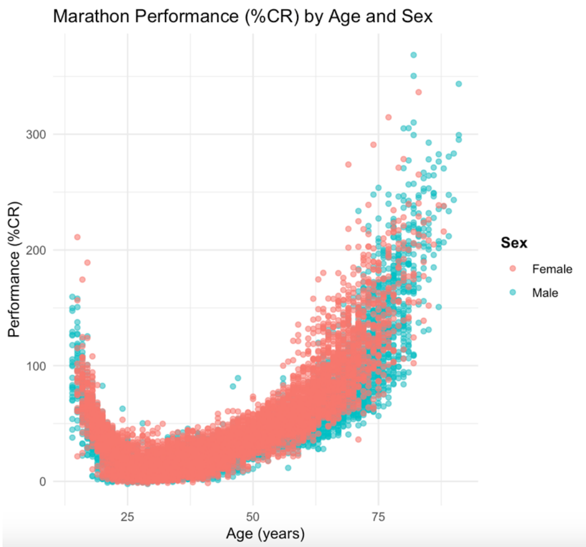

Figure 1: Scatter plot illustrating marathon performance (%CR) across different age groups and sexes, showing a U-shaped performance trend with age.

The scatter plot offers a detailed visualization of how marathon performance, represented as a percentage of a reference value (%CR), varies across different age groups and between sexes. The plot reveals a clear U-shaped curve, demonstrating that performance declines from young adulthood into middle age, then decrease more in later years. This trend is likely influenced by physiological changes that come with aging, such as decreased muscle mass and aerobic capacity, which initially cause a dip in performance. Over time, this decline is tempered as only the more experienced or healthier older adults may continue participating in marathons.

The plot also shows that both males and females generally follow the same performance trajectory, although there's a notable difference in the scatter and density of data points. Older female runners show a higher variability in performance compared to males, suggesting that factors such as training intensity, injury rates, or participation levels, which might differ between genders, can affect individual performances. Furthermore, the density of data points reduces significantly with age, especially beyond the 70-year mark, indicating fewer participants in these older age groups. The high concentration of data points between ages 30 and 60, however, might influence the modeling accuracy and the perceived performance trends within this age range, potentially emphasizing trends that are more representative of this demographic.

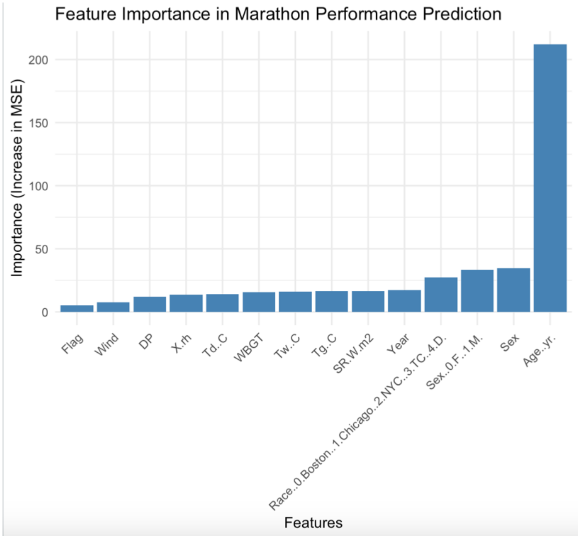

Figure 2: Feature importance plot from a Random Forest model, showing age as the most significant predictor of marathon performance, alongside factors like sex, year, and marathon locations.

In the feature importance plot from the Random Forest model, age emerges as the most significant predictor of marathon performance. This supports the observed trends from the scatter plot and highlights the strong impact of age-related changes on performance. The model also points out other influential factors such as the interaction of sex and age, specific years, and the locations of marathons like Boston, NYC, and Chicago, indicating that both physiological and demographic factors along with temporal and spatial variations significantly impact marathon outcomes.

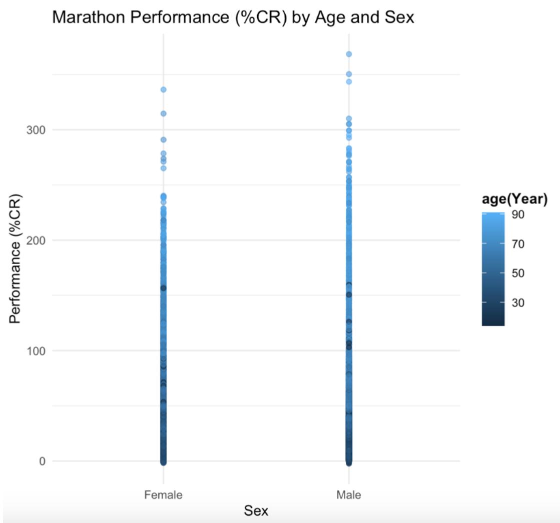

Figure 3: Scatter plot displaying marathon performance (%CR) by sex, with points colored by age, highlighting differences in performance trends between male and female runners across various age groups.

Lastly, the scatter plot separating marathon performance by sex shows how performances are distributed across different age and within each sex. This plot illustrates that peak performances typically occur in the middle age range for both sexes, supporting our hypothesis that age factors affect marathon running across genders similarly. The color gradient used in this plot effectively highlights the variation in performance across the lifespan, emphasizing that while sex does have an impact, age is a more dominant factor, with younger and older runners generally exhibiting lower performance levels compared to their middle-aged counterparts.

AIM 2: Environmental Impact

The regression analysis presented in the figures deeply explores the impact of Wet Bulb Globe Temperature on marathon performance, integrating variables such as age and sex to underscore how these factors influence athletes' responses to environmental conditions. WBGT, a measure that combines temperature, humidity, wind speed, and solar radiation to provide a comprehensive index of heat stress, is shown to have a significant negative correlation with performance.

Figure 4: Line graph illustrating the decline in marathon performance as WBGT increases, with separate trend lines for males and females.

Figure 4 vividly demonstrates how marathon performance degrades as WBGT increases, with the performance drop being more severe for older adults and males. This could indicate less efficient thermoregulatory responses in these groups, potentially due to physiological declines in heat adaptation capabilities with age and possible sex-related differences in body heat management. The plotted lines for each sex show a clear downward trajectory as WBGT rises, with the shaded areas possibly representing confidence intervals that highlight the variability and certainty of the predictions. The trend is starker for males, suggesting that they might be more vulnerable to higher temperatures compared to females, a factor that could be vital for trainers and event organizers to consider when planning training schedules and race logistics.

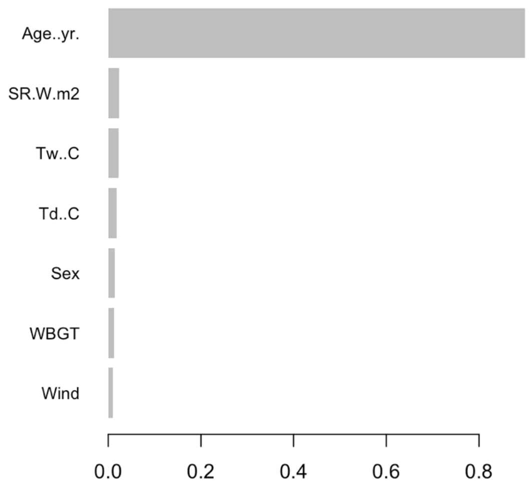

Figure 5: Scatter plot showing individual marathon performances against WBGT values, illustrating the negative impact of higher WBGT on performance and variability among athletes.

	Feature	Gain	Cover	Frequency
	<char>	<num>	<num>	<num>
1:	Age..yr.	0.899551330	0.29419988	0.21337019
2:	SR.W.m2	0.023388719	0.16697083	0.17165117
3:	Tw..C	0.022091475	0.13171907	0.16813270
4:	Td..C	0.018491914	0.14213565	0.19301332
5:	Sex	0.014215908	0.05552480	0.02488062
6:	WBGT	0.012877018	0.11231352	0.11409902
7:	Wind	0.009383634	0.09713626	0.11485298

Figure 5 plots individual performances against WBGT values by fitting a Xgboost model to illustrate the feature importance. This visualization not only confirms the general negative impact of higher WBGT on performance but also tells how individual performances vary under different conditions. In the broader context of marathon preparation and safety, these insights are critical. They suggest that maintaining lower WBGT levels during races, possibly through scheduling events during cooler parts of the day or in seasons with milder weather, could significantly enhance overall performance and athlete safety. For training, these findings underscore the importance of heat acclimatization programs, particularly for older runners and males who are shown to be more adversely affected by higher WBGT levels.

Additionally, the use of modeling techniques like XGBoost provides a robust framework for predicting the impact of various factors on marathon performance. By training the model with parameters that account for non-linear relationships and interaction effects among features like age, sex, and environmental variables, we can derive an understanding of how these elements combine to affect performance. The feature importance plot generated from the model would further clarify which variables are most critical in predicting performance, enabling focused interventions and informed decision-making for athletes and coaches.

AIM 3: Weather Parameter Importance

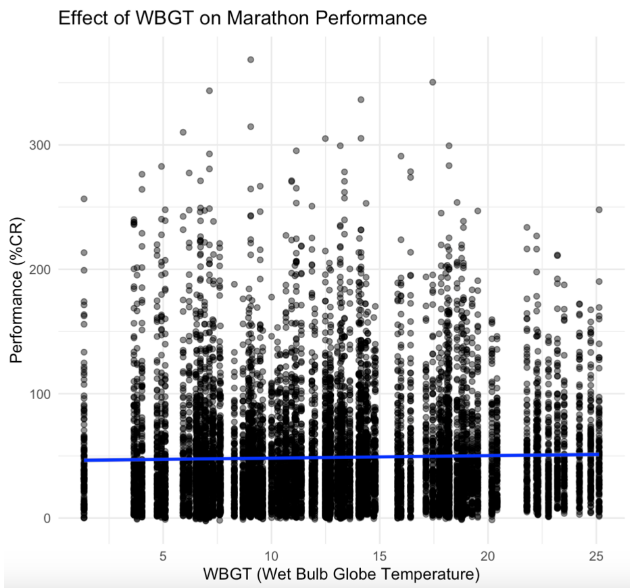

Figure 6: Scatter plot illustrating the relationship between marathon performance (%CR) and Wet Bulb Globe Temperature (WBGT)

The scatter plot provides a dense aggregation of performance data against the Wet Bulb Globe Temperature (WBGT), a comprehensive metric for assessing environmental heat stress. This visualization offers a panoramic view of how varying levels of WBGT correlate with marathon performance, measured as a percentage of a reference standard (%CR).

In this plot, each data point represents an individual marathon performance under different WBGT conditions. The horizontal blue line likely represents the average or a significant threshold in marathon performance across varying temperatures. From the visual, it's evident that most of the data points are clustered around lower WBGT values, suggesting better performance outcomes in cooler conditions. As WBGT increases, the distribution of performance points becomes sparser and variable, indicating a decrease in performance as conditions become hotter. This is consistent with what we know about the physiological impacts of heat on endurance sports: higher temperatures exacerbate dehydration, reduce heat dissipation, and can increase core body temperature, all of which impair physical performance. The inclusion of demographic factors like age and sex alongside WBGT in the feature importance model underscores the multifaceted nature of performance determinants, emphasizing an intricate interplay between human physiology and environmental conditions. This relationship highlights that while environmental conditions are critical, the demographic characteristics of the runners also play a crucial role in how these environmental stressors affect performance.

This scatter plot is not only visualizing the direct correlation between WBGT and marathon performance but also sets the stage for deeper analytical pursuits. By establishing the foundational impacts of WBGT visually, we can better appreciate the complex dynamics that feature importance models aim to unravel, enhancing our understanding of how to optimize training and competitive strategies in marathons to mitigate the adverse effects of high environmental heat stress. This insight is particularly valuable for coaches, sports scientists, and event organizers aiming to maximize athlete safety and performance through strategic planning and personalized training adjustments.

5. Discussion

The detailed analysis underscores the complexity of marathon performance as a multifaceted phenomenon influenced by age, sex, and environmental factors, particularly Wet Bulb Globe Temperature (WBGT). It confirms that aging brings about physiological changes that can adversely affect marathon performance, with a noted acceleration in performance degradation among female’s post-menopause. This may be attributed to physiological and hormonal changes that impact muscle mass, strength retention, and cardiovascular efficiency, aligning with existing literature that highlights the physiological limitations imposed by aging.

The impact of environmental conditions on performance points to the need of advanced strategies for older athletes and those in warmer climates. The significant influence of WBGT, which combines temperature, humidity, solar radiation, and wind speed, suggests that heat is a critical factor in endurance sports. It tells us that managing heat exposure is crucial, particularly for older athletes who may have diminished thermoregulatory responses. Furthermore, the observed sex differences in the trajectory of performance decline underline the need for personalized training and competition strategies that consider physiological and hormonal distinctions between genders. For instance, strategies that bolster heat acclimatization may benefit older female athletes, helping to mitigate the effects of reduced estrogen levels on thermoregulation.

In addition, the findings enhance the understanding that trainers and organizers need to look beyond average temperatures in their logistical planning. Comprehensive weather assessments that include measures like WBGT are essential for optimizing both performance and safety. This approach not only aids in better race day planning but also in training regimens that prepare athletes for the specific conditions they will face during competitions. From a public health perspective, these results are invaluable. They could inform advisories and guidelines that regulate or advise on outdoor activities under various climatic conditions, especially for vulnerable groups such as the elderly. Public health strategies could include recommendations for the best times of day for physical activity or warnings during extreme weather conditions to prevent heat-related illnesses.

Future research should continue to explore the interactions among age, sex, and environmental factors in marathon performance. A deeper understanding could lead to breakthroughs in sports science that prioritize athlete care and optimize performance through innovative training techniques, adaptive clothing technologies, and personalized hydration strategies. Such research might also explore genetic factors that could predispose certain athletes to better handle environmental stresses, opening the door to even more tailored approaches in athletic training and competition planning. This ongoing investigation will undoubtedly refine our strategies and enhance the safety and efficacy of training programs for marathon runners across different demographics and environments.

Summary

In summary, AIM 1 aims to find the effects of increasing age on marathon performance in both men and women. Our findings revealed a clear decline in performance with age, with a decrease observed in female athletes' post-menopause. This suggests that physiological and hormonal changes due to aging significantly impact performance, emphasizing the need for age-specific and gender-specific training strategies to mitigate these effects.

For AIM 2, we focused on exploring the impact of environmental conditions, particularly temperature and humidity as encapsulated by Wet Bulb Globe

Temperature on marathon performance. The study found that higher WBGT levels significantly decrease performance, with older athletes and males experiencing more substantial effects. This underscores the importance of personalized acclimatization strategies and the need for marathon events to consider these environmental impacts when planning and conducting races to ensure athlete safety and optimal performance.

Finally, AIM 3 aimed to identify the specific weather parameters, like WBGT, temperature, and other flag conditions, that have the largest impact on marathon performance. The analysis confirmed that WBGT is a critical factor, affecting performance more significantly than other measured parameters. This highlights the necessity for marathon organizers and trainers to integrate comprehensive weather assessments into their planning processes to mitigate the adverse effects of environmental conditions on athletes.

In conclusion, the study bridges gaps of how age, sex, and environmental factors and marathon performance. It informs athletic trainings, optimizes event scheduling, and assist the public health guidelines, particularly considering global climate changes affecting outdoor sports and activities. This research not only enhances our understanding but also offers practical approaches to promoting health and performance in endurance sports among changing environmental conditions.

Reference

1. Knechtle, Beat, et al. "Trends in Weather Conditions and Performance by Age Groups Over the History of the Berlin Marathon." *Frontiers in Physiology*, vol. 12, 12 May 2021, doi:10.3389/fphys.2021.654544.
2. **Knechtle, Beat, et al.** "The Role of Environmental Conditions on Master Marathon Running Performance in 1,280,557 Finishers the 'New York City Marathon' From 1970 to 2019." *Frontiers in Physiology*, vol. 12, 16 May 2021, https://doi.org/10.3389/fphys.2021.665761.
3. **Weiss, Katja, et al.** "The Influence of Environmental Conditions on Pacing in Age Group Marathoners Competing in the 'New York City Marathon'." *Frontiers in Physiology*, vol. 13, 13 June 2022, https://doi.org/10.3389/fphys.2022.842935.
4. **Nikolaidis, Pantelis T., et al.** "The Role of Environmental Conditions on Marathon Running Performance in Men Competing in Boston Marathon from 1897 to 2018." *International Journal of Environmental Research and Public Health*, vol. 16, no. 4, 2019, https://doi.org/10.3390/ijerph16040614.
5. **Knechtle, Beat, et al.** "The Role of Weather Conditions on Running Performance in the Boston Marathon from 1972 to 2018." *PLoS ONE*, 8 Mar. 2019, https://doi.org/10.1371/journal.pone.0212797.

code appendix

```
library(ggplot2)

library(randomForest)

library(dplyr)

library(xgboost)


marathon_data <- read.csv("project1.csv")


print(head(data))

summary(marathon_data)


print(marathon_data)


# AIM 1: Scatter plot illustrating marathon performance (%CR) across different age
groups and sexes

marathon_data$Sex <- factor(marathon_data$Sex..0.F..1.M., labels = c("Female",
"Male"))


ggplot(marathon_data, aes(x=Age..yr., y=X.CR, color=Sex)) +

  geom_point(alpha=0.6) +

  labs(title = "Marathon Performance (%CR) by Age and Sex",

       x = "Age (years)",

       y = "Performance (%CR)",

       color = "Sex") +

  theme_minimal() +

  theme(legend.title = element_text(face = "bold"))
```

```
## AIM 1: Scatter plot displaying marathon performance (%CR) by sex, with points colored by age
marathon_data$Sex <- factor(marathon_data$Sex..0.F..1.M., labels = c("Female", "Male"))

ggplot(marathon_data, aes(x=Sex, y=X.CR, color=Age..yr.)) +
  geom_point(alpha=0.6) +
  labs(title = "Marathon Performance (%CR) by Age and Sex",
       x = "Sex",
       y = "Performance (%CR)",
       color = "age(Year)") +
  theme_minimal() +
  theme(legend.title = element_text(face = "bold"))


### AIM 1: Feature importance plot from a Random Forest model
marathon_data <- na.omit(marathon_data)
sum(is.na(marathon_data$X.CR))   # This should return 0

# Fit the Random Forest model
set.seed(123)
rf_model <- randomForest(X.CR ~ ., data=marathon_data, ntree=500, importance=TRUE)

importance_data <- importance(rf_model, type=1)
feature_importance <- data.frame(Feature = rownames(importance_data), Importance = importance_data[, "%IncMSE"])
```

```
ggplot(feature_importance, aes(x=reorder(Feature, Importance), y=Importance)) +
  geom_col(fill="steelblue") +
  labs(title="Feature Importance in Marathon Performance Prediction",
       x="Features",
       y="Importance (Increase in MSE)") +
  theme_minimal() +
  theme(axis.text.x=element_text(angle=45, hjust=1))



# AIM 2: Line graph illustrating the decline in marathon performance as WBGT increases
library(ggplot2)
library(dplyr)
library(lme4)

marathon_data$Sex <- factor(marathon_data$Sex..0.F..1.M., labels = c("Female", "Male"))
model <- lm(X.CR ~ Age..yr. * Sex + Td..C + Tw..C + Wind + WBGT + SR.W.m2, data=marathon_data)
summary(model)

interaction_plot <- marathon_data %>%
  group_by(Age..yr., Sex) %>%
  summarise(Avg_CR = mean(X.CR, na.rm = TRUE), Avg_WBGT = mean(WBGT, na.rm = TRUE)) %>%
  ggplot(aes(x=Avg_WBGT, y=Avg_CR, color=Sex)) +
```

```
  geom_point() +
  geom_smooth(method="lm") +
  labs(title="Impact of WBGT on Marathon Performance by Age and Sex",
       x="Average WBGT",
       y="Average Performance (%CR)",
       color="Sex") +
  theme_minimal()

print(interaction_plot)



## AIM2: Feature importance
marathon_data$Sex <- as.numeric(factor(marathon_data$Sex..0.F..1.M.))


data_matrix <- xgb.DMatrix(data = as.matrix(marathon_data[, c("Age..yr.", "Sex",
"Td..C", "Tw..C", "Wind", "WBGT", "SR.W.m2")]),
                           label = marathon_data$X.CR)


params <- list(
  booster = "gbtree",
  objective = "reg:squarederror",
  eta = 0.1,
  gamma = 0,
  max_depth = 6,
  min_child_weight = 1,
  subsample = 0.5,
  colsample_bytree = 0.5
```

```
)

set.seed(123)
xgb_model <- xgb.train(params = params, data = data_matrix, nrounds = 100, verbose = 0)

importance_matrix <- xgb.importance(feature_names = colnames(data_matrix), model = xgb_model)
print(importance_matrix)
xgb.plot.importance(importance_matrix)


## AIM3: Identify the weather parameters (WBGT, Flag conditions, temperature, etc) that have the largest impact on marathon performance
library(ggplot2)
library(dplyr)
library(corrplot)

correlation_matrix <- cor(marathon_data[, c('X.CR', 'WBGT', 'Td..C', 'Tw..C', 'Wind', 'SR.W.m2')], use = "complete.obs")
corrplot(correlation_matrix, method = "circle")

model <- lm(X.CR ~ WBGT + Td..C + Tw..C + Wind + SR.W.m2, data=marathon_data)
summary(model)

ggplot(marathon_data, aes(x=WBGT, y=X.CR)) +
  geom_point(alpha=0.5) +
```

```
geom_smooth(method="lm", color="blue") +
labs(title="Effect of WBGT on Marathon Performance",
     x="WBGT (Wet Bulb Globe Temperature)",
     y="Performance (%CR)") +
theme_minimal()
```