

# Exploring the Relationships between Environmental Conditions and Marathon Performance

Yunan Chen

2024-10-06

## Abstract

Environmental conditions are critical for endurance sports such as marathons and may affect a runner's performance. The effects may vary by gender and age. Utilizing data from five marathons collected over a 17 to 23-year period, this report explores the relationship between environmental variables and marathon performance across gender and age. The findings suggest a "U-shape" relationship between age and performance. Weather conditions were more likely to affect performance in the elderly than in the youth and adults. Limitations such as data coding errors, missing data, lack of information on extreme weather conditions, and the non-linear relationship emphasize the need for further research and examination.

## Introduction

The influence of weather conditions on athletic performance, especially in endurance activities such as marathons, has raised considerable attention in recent years. Marathon running presents an extraordinary physical challenge, requiring sustained performance over a long distance, which is heavily influenced by the body's ability to regulate temperature and adapt to environmental factors. Among weather-related variables, temperature (both heat and cold), relative humidity, wind speed, solar radiation, and air quality have been identified as key contributors to performance outcomes. Research by Ely et al. (2007) found that higher temperatures can significantly slow marathon runners, with the impact worsening as temperatures increase. Older adults face additional challenges due to age-related declines in physiological functions like reduced sweating, diminished skin blood flow, and lower heart function, making them less able to handle heat or conserve warmth in cold conditions (Kenney & Munce, 2003). Sex differences also play a role in how runners handle weather, as women's higher surface area-to-mass ratio can help them lose heat better in hot and humid environments, but their lower sweat production compared to men may limit this benefit. Conversely, men's greater sweating capacity gives them an advantage in dry heat (Yanovich et al., 2020). Understanding these differences is essential for exploring how weather impacts marathon performance across age groups and genders, helping athletes prepare and minimize risks during races.

This report, in collaboration with Dr. Brett Romano Ely and Dr. Matthew Ely from the Department of Health Sciences at Providence College, aims to explore the interactions between age, sex, and environmental conditions in influencing marathon performance. It investigates the specific effects of factors such as temperature, humidity, solar radiation, and wind, while examining how these variables uniquely impact runners across different ages and genders. By identifying the weather conditions with the greatest impact, this analysis seeks to inform preparation strategies and enhance safety measures for a diverse range of marathon participants.

## Data Collection

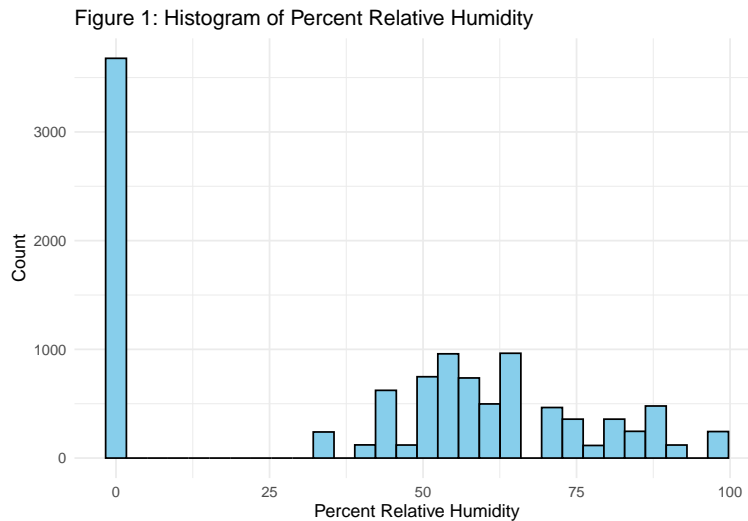
Data used in this report was provided by Dr. Brett Romano Ely and Dr. Matthew Ely. The primary dataset includes top single-age performances from five major marathons over 17–23 years for both men and women,

along with detailed weather conditions for each marathon. The performance variable, **%CR**, represents the percentage by which a participant's performance deviates from the current course record. A lower **%CR** value indicates a performance closer to the course record. Wet Bulb Globe Temperature (WBGT) was calculated using the formula  $WBGT = 0.7 \cdot \text{Wet Bulb Globe Temperature} + 0.2 \cdot \text{Black Bulb Globe Temperature} + 0.1 \cdot \text{White Bulb Globe Temperature}$ , and the variable **Flag** was grouped into five categories based on the value of **WBGT** and risk of heat illness, the temperature-related variables in the data are expected to be highly correlated.

The other two datasets provide information on air quality and course records, respectively. The AQI dataset, however, required modifications before merging with the primary dataset. This dataset contains detailed air quality measurements across multiple sites, dates, and durations (e.g., 1-hour, 8-hour, and 24-hour averages), including variables such as local dates, county codes, and marathon names. To align the AQI data with the primary dataset, it was summarized to focus on broader air quality trends relevant to marathon events, which typically span several hours and cover large areas. Hourly measurements were excluded to minimize unnecessary variability, while longer-duration averages (e.g., 8-hour and 24-hour averages) provided a more stable and representative measure of air quality. The `ymd()` function from the `lubridate` package was used to standardize the local date variable into a consistent format and extract the year, which was essential for summarizing the AQI data. The summarized data includes the average AQI for each marathon and its corresponding year. The modified AQI dataset and the course record dataset were then merged with the primary dataset by Year and Marathon, enabling a comprehensive analysis of the combined data.

## Data Pre-processing

The raw dataset `project 1` contained 11,564 observations and 14 variables. To get a more comprehensive dataset, the average air quality information of each marathon was calculated using the variable **AQI** in dataset `aqi_values` and was merged to the major data set `project 1` by years and marathon. Some values of the humidity variable appeared to be reported incorrectly. By looking at the histogram in percent relative humidity, we can see that more than 3500 observations fall around 0, which is highly unlikely to be the case since the variable should be in percentage (**Figure 1**). Therefore, we modified the values that were less than or equal to 1 by multiplying them by 100. In addition, in the original dataset, missing values in the categorical variable **Flag** were coded as empty cells, which would cause R to interpret them as one of the categories by default. Therefore, these values were recoded as NA.



The chosen age cutoffs are supported by research findings and reflect meaningful shifts in physical capabilities relevant to marathon performance. Research suggests that marathon performance typically peaks between the ages of 25 and 35, with a noticeable decline starting around age 35. This decline becomes more evident

in the 40–54 age group, as runners in this range often experience the cumulative effects of aging, including reduced maximal oxygen uptake (VO2 max), slower recovery times, and decreased muscle mass, all of which impact endurance and speed. The decline continues into the 55–69 age group, where physiological changes such as further reductions in VO2 max, increased susceptibility to injury, and a general slowing of metabolic processes significantly affect performance. The sharp declines observed in runners aged  $\geq 70$  years further justify this categorization, as performance in this group is typically marked by advanced aging effects. These cutoffs align with well-documented life stages in endurance capabilities (Zavorsky et al., 2017). Additionally, these age cutoffs create a balanced distribution of participants, ensuring reasonable proportions in each group for analysis. As shown in Table 1, the  $<25$  years group and the  $\geq 70$  years group both contain around 14% of participants, while the 25–39 years group, 40–54 years group, and 55–69 years group each contain approximately 24% of participants. Smaller proportions in the youngest and oldest age groups and larger proportions in the middle categories reflect the typical demographic distribution of marathon participants, as younger and older runners are less common compared to those in their peak and middle-age years. These distributions demonstrate that the cutoffs effectively capture a broad range of participants and maintain sufficient representation across all age groups.

Table 1: Number of Participants by Age Group and Gender

Characteristic	Gender	
	Female N = 5,452	Male N = 6,112
<b>Age Group</b>		
< 25 yrs	788	834
25-39 yrs	1,440	1,440
40-54 yrs	1,440	1,440
55-69 yrs	1,346	1,435
$\geq 70$ yrs	438	963

Table 2 shows all weather-related information and the course record by marathon. We can see that data were collected on five marathons: Boston, Chicago, New York City, Twin Cities, and Grandma’s Marathon. The “N” next to the marathon’s name stands for the number of years the data collection crossed. For example, data on the Boston Marathon were collected across 18 years. Information on weather conditions in Chicago, New York, the Twin Cities, and Grandma’s Marathon was missing for one of the years. The 2011 Chicago, New York and Twin Cities Marathons and the 2012 Grandma’s Marathon appeared to be missing all weather-related variables. Further research may be needed to determine why this happened.

Table 2: Summary Statistics of Weather Conditions by Race

Characteristic	Race				
	Boston Marathon N = 18	Chicago Marathon N = 21	Grandma’s Marathon N = 17	New York City Marathon N = 23	Twin Cities Marathon N = 17
<b>Flag</b>					
WBGT <10C	9 (50%)	6 (29%)	0 (0%)	11 (48%)	5 (29%)
WBGT 10-18C	7 (39%)	12 (57%)	6 (35%)	7 (30%)	7 (41%)
WBGT >18-23C	1 (5.6%)	1 (4.8%)	8 (47%)	4 (17%)	3 (18%)
WBGT >23-28C	1 (5.6%)	1 (4.8%)	2 (12%)	0 (0%)	1 (5.9%)
Missing	0 (0%)	1 (4.8%)	1 (5.9%)	1 (4.3%)	1 (5.9%)
<b>Dry bulb temperature</b>	11.6 (6.0)	12.4 (6.2)	18.9 (3.4)	11.7 (4.8)	13.2 (5.7)

Table 2: Summary Statistics of Weather Conditions by Race (*continued*)

Characteristic	Race				
	Boston Marathon N = 18	Chicago Marathon N = 21	Grandma's Marathon N = 17	New York City Marathon N = 23	Twickenham Marathon N = 17
Missing	0	1	1	1	1
<b>Wet bulb temperature</b>	7.6 (3.9)	8.6 (5.9)	14.9 (2.5)	7.6 (5.1)	9.9 (5.6)
Missing	0	1	1	1	1
<b>Percent relative humidity</b>	61 (21)	61 (11)	68 (16)	55 (18)	64 (16)
Missing	0	1	1	1	1
<b>Black globe temperature</b>	24 (9)	25 (6)	32 (8)	21 (6)	25 (7)
Missing	0	1	1	1	1
<b>Solar radiation in Watts</b>	654 (191)	460 (96)	679 (195)	401 (134)	437 (143)
Missing	0	1	1	1	1
<b>Wind</b>	12.0 (4.6)	8.2 (3.3)	9.2 (2.9)	11.2 (4.7)	8.8 (3.3)
Missing	0	1	1	1	1
<b>Dew Point</b>	3 (5)	5 (7)	12 (3)	3 (7)	6 (8)
Missing	0	1	1	1	1
<b>Course Record in seconds</b>	8,445 (47)	8,312 (111)	8,849 (60)	8,611 (64)	8,819 (23)
<b>Average Air Quality</b>	42 (15)	40 (13)	37 (15)	33 (14)	35 (15)

<sup>1</sup> Mean (SD) for continuous; n (%) for categorical

## Marathon Performance by Age and Gender

We first examined the effects of increasing age on marathon performance in both men and women. The analysis began by exploring the relationship between age and performance using age as a continuous variable, followed by an examination using age groups. The minimum age was 14 years, and the maximum age was 91 years.

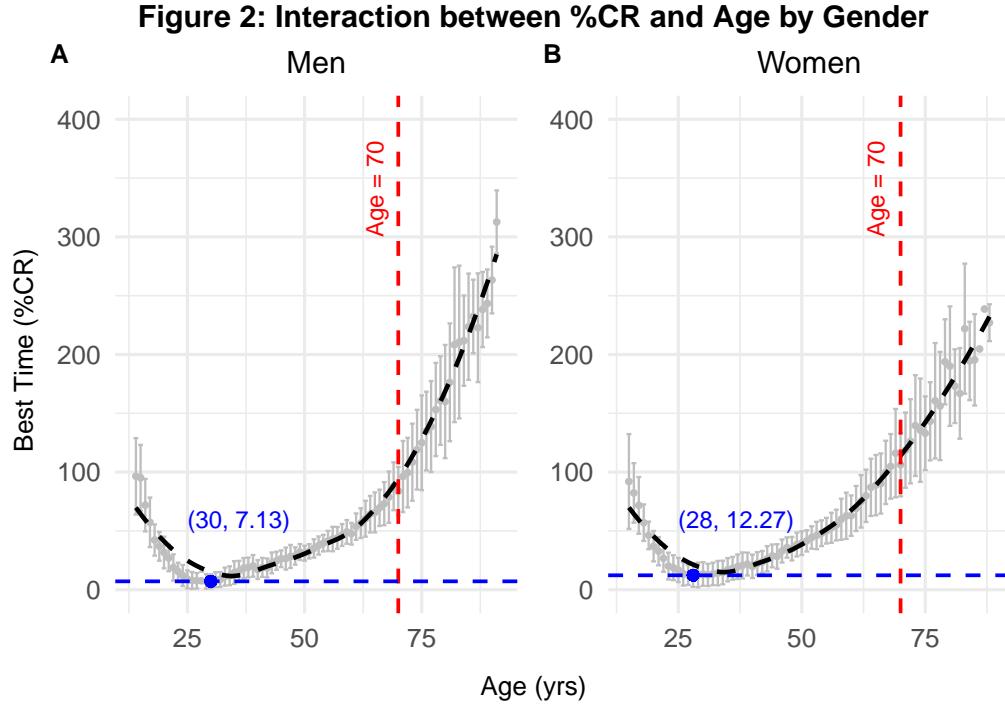
Figure 2 shows the interaction between %CR (percent off the current course record) and age for men (panel A) and women (panel B), highlighting a clear “U-shaped” non-linear relationship. This suggests the need for a quadratic term for age when fitting the model. Performance, as measured by %CR, improves steadily during younger ages, with runners gaining speed and efficiency as they progress from adolescence (ages 15–19) into early adulthood (ages 20–29). However, these improvements slow in the late 20s and early 30s, marking the peak performance years. The blue dots in the plot indicate the ages of peak performance: age 30 for men (7.13 %CR) and age 28 for women (12.27 %CR). After the peak, as age increases beyond 30, %CR steadily rises, indicating a gradual decline in performance. Between ages 30 and 70, men’s %CR increases from 7 to 83, representing an increase of 76, while women’s %CR increases from 12 to 106, reflecting an increase of 94. This increase in %CR, or decline in performance, reflects natural aging processes, with women showing a larger decline in this age range.

After age 70, the decline in performance accelerates sharply. Men’s %CR increases from 83 to over 300, while women’s %CR rises from 106 to over 300, with men showing a larger decline during this phase. This suggests that age has a greater impact on men’s performance in their later years compared to women’s. The red dashed line at age 70 marks this turning point, where the steep decline highlights the effects of

Table 3: Summary Statistics of %CR by Gender, Mean (Q1, Q3)

Characteristic	Men, N=6,112	Women, N=5,452
CR% (by Age Group)		
< 25 yrs	36.12 (16.9, 44.77)	40.84 (23.93, 52.95)
25-39 yrs	11.64 (4.42, 17.62)	15.55 (6.58, 22.97)
40-54 yrs	27.99 (21.69, 34.6)	34.06 (24.85, 43.07)
55-69 yrs	58.75 (44.36, 68.37)	75.35 (54.47, 89.25)
$\geq 70$ yrs	132.48 (90.13, 164.09)	138.61 (103.96, 166.44)

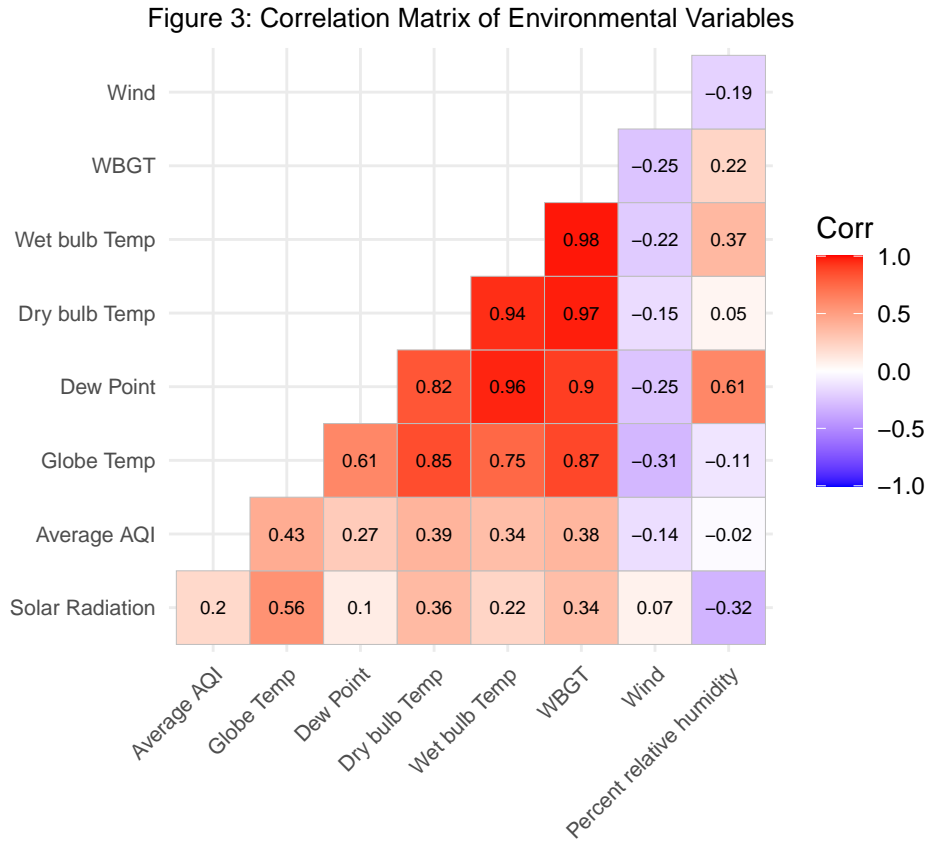
advanced aging, including significant losses in endurance and strength. The widening confidence intervals in both panels reflect greater variability in performance among older runners, likely due to differences in health, training history, and adaptation to aging.



To further explore whether the patterns observed in Figure 2 hold when using age as a categorical variable, Table 3 was created to compare the mean %CR across age groups and genders. For both men and women, the lowest average %CR was found in the 25–39 age group, with a mean of 11.64 for men and 15.55 for women. This indicates that men and women in the 25–39 age group had, on average, the best performance, with times closest to the course record. In contrast, participants in the oldest age group ( $\geq 70$  years) had a mean %CR exceeding 100%, showing that older participants ran more than twice as slow as the course record. Across all age groups, men generally performed slightly better than women, with men’s %CR consistently 4 to 7 points lower than women’s. The trends in performance were similar for men and women across all age groups, showing a decline in performance both below and above the 25–39 age group, reinforcing the “U-shaped” pattern previously observed in Figure 2.

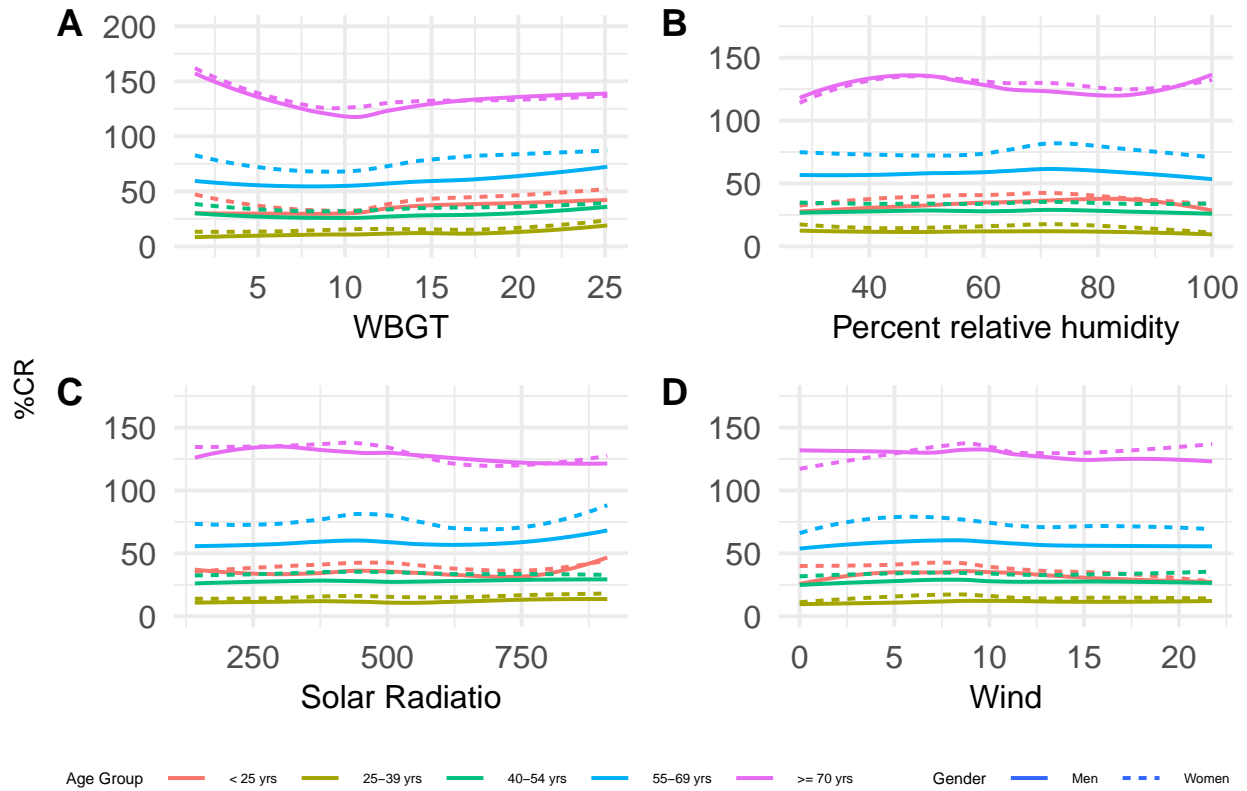
## Potential Enviornmentmental Variables

Next, we examine the effects of environmental conditions on marathon performance in both males and females. Figure 3 presents the correlation matrix of environmental variables, guiding the selection of WBGT, humidity, solar radiation, and wind for the analysis. WBGT was chosen as a key variable because it effectively combines multiple components of heat stress into a single measure, as defined by the formula:  $WBGT = 0.7 \cdot \text{Wet Bulb Globe Temperature} + 0.2 \cdot \text{Black Bulb Globe Temperature} + 0.1 \cdot \text{Globe Temperature}$ . Given this definition and the high correlations between WBGT, wet bulb temperature, black bulb temperature, and globe temperature (as seen in the plot), we excluded the latter variables to avoid redundancy. Additionally, humidity, solar radiation, and wind were included as they provide distinct environmental information not fully encompassed by WBGT, ensuring the exploration considers the unique contributions of these factors.



**Figure 4** illustrates the relationship between marathon performance, measured as percentage off the current course record (%CR), and environmental factors: Wet Bulb Globe Temperature (WBGT, panel A), percent relative humidity (panel B), solar radiation (panel C), and wind (panel D). The plots are stratified by age group (<25 years, 25–39 years, 40–54 years, 55–69 years, and  $\geq 70$  years), indicated by different colors, and by gender (solid lines for men and dashed lines for women).

**Figure 4: Environmental Factors and Marathon Performance**



## WBGT

Wet Bulb Globe Temperature (WBGT) is a composite measure of temperature that accounts for dry, wet, and globe temperatures, providing an indication of heat stress in the environment. Higher WBGT values reflect greater environmental heat stress, making it more difficult for the body to regulate internal temperature. This increases the physical demands and risks associated with activities such as running a marathon. **Figure 4 (Panel A)** visualizes the relationship between WBGT and marathon performance (%CR) across different age groups for men and women, with performance trends represented by smoothed lines. The lines are color-coded by age group (<25 years, 25–39 years, 40–54 years, 55–69 years, and  $\geq 70$  years), and line types differentiate gender (solid for men and dashed for women). The fitted lines reveal that, regardless of WBGT levels, the 25–39 age group consistently outperforms all other age groups, while the  $\geq 70$  age group performs the worst. This pattern is consistent for both men and women, highlighting the superior endurance and performance capabilities of the 25–39 age group compared to the significant challenges faced by the oldest group. Performance (%CR) exhibits a distinct non-linear relationship with WBGT across all age groups. When WBGT is below 10, %CR decreases (indicating better performance) as WBGT increases, suggesting slightly warmer conditions may improve performance. However, once WBGT exceeds 10, %CR rises sharply (indicating worse performance), reflecting the detrimental effects of extreme heat stress. This rise and fall in %CR is most pronounced in the  $\geq 70$  age group, which shows the greatest sensitivity to WBGT changes. In contrast, the flatter lines observed for younger age groups indicate their performance is less affected by changes in WBGT, demonstrating greater resilience to environmental heat stress. Overall, **Panel A** underscores the importance of WBGT as a key environmental factor influencing marathon performance, particularly for older runners who are more vulnerable to heat stress.

## Relative humidity

Relative humidity is a measure of the amount of moisture in the air relative to its capacity to hold water vapor. High levels of humidity can make physical activity more challenging by reducing the body's ability to cool through sweat evaporation. **Figure 4 (Panel B)** illustrates the relationship between average relative humidity and marathon performance (%CR) across different age groups for men and women. The oldest age group ( $\geq 70$  years) exhibited the most pronounced fluctuations in performance, as indicated by the curviest fitted lines. For this group, %CR initially increased (indicating worse performance) when humidity rose from 0 to 50%, remained relatively stable between 50% and 80%, and then increased again when humidity exceeded 80%. The 55–69 age group displayed a similar pattern, though the variations were less pronounced compared to the  $\geq 70$  group. Across all age groups, performance was relatively stable when humidity ranged between 50% and 80%, suggesting that moderate humidity has minimal impact on marathon performance. For younger age groups (<25, 25–39, and 40–54 years), the fitted lines were much flatter, indicating that these participants were less affected by changes in humidity. The consistent trends between men (solid lines) and women (dashed lines) suggest that the relationship between humidity and performance is similar for both genders, with no notable differences observed. Overall, Panel B highlights that older runners, particularly those in the  $\geq 70$  age group, are more sensitive to changes in humidity, while younger runners show greater resilience to its effects. This suggests that age plays a significant role in determining how humidity impacts marathon performance.

## Solar radiation

Solar radiation refers to the energy emitted by the sun, which contributes to environmental heat stress during outdoor activities. High levels of solar radiation can increase the risk of overheating and fatigue in runners. **Figure 4 (Panel C)** illustrates the relationship between solar radiation and marathon performance (%CR) across different age groups for men and women. The position of the fitted lines shows that participants in the 25–39 age group consistently performed the best, regardless of solar radiation levels, followed by the 40–54, <25, 55–69, and  $\geq 70$  age groups. For both men and women in the <25, 25–39, and 40–54 age groups, the fitted lines are relatively flat when solar radiation is below 500, indicating that solar radiation in this range has little to no effect on their performance. However, a distinct “U-shaped” relationship is observed for females in the 55–69 and  $\geq 70$  age groups. For these groups, performance (%CR) initially worsens with increasing solar radiation but begins to improve after solar radiation exceeds 500. This suggests that older female runners are more sensitive to changes in solar radiation, with their performance declining at moderate levels but potentially adapting or stabilizing at higher levels. In contrast, younger age groups appear largely unaffected by variations in solar radiation, demonstrating greater resilience to this environmental factor.

## Wind

Wind speed measures the movement of air in the environment, which can either assist or hinder a runner's performance depending on its direction and intensity. **Figure 4 (Panel D)** illustrates the relationship between wind speed and marathon performance (%CR) across different age groups for men and women. The fitted lines show that participants in the 25–39 age group consistently performed the best, while those in the  $\geq 70$  age group performed the worst on average, regardless of wind speed. For the oldest age group ( $\geq 70$  years), the fitted line is more pronouncedly curvy for females compared to males, suggesting that wind speed introduces greater variability in performance for older women. Specifically, for females in this group, %CR increased (indicating worse performance) as wind speed rose from 0 to 7.5, while no such trend was observed for males in the same age group. This highlights a potential gender difference in sensitivity to wind conditions among older runners. For younger age groups (<25, 25–39, and 40–54 years), the fitted lines remained relatively flat, indicating that wind speed had minimal impact on marathon performance for both sexes. These findings suggest that the influence of wind speed on performance is more prominent among older participants, particularly older women, while younger runners are largely unaffected by changes in wind conditions.



## Importance Variables

To quantify the association between weather-related variables and marathon performance, a quadratic linear model was used. The model includes the covariates Wet Bulb Globe Temperature (WBGT), relative humidity, solar radiation, wind, and average air quality (AQI), as well as their interactions with both the linear and quadratic terms of age. The inclusion of a quadratic age term ( $\text{Age}^2$ ) accounts for the observed non-linear relationship between age and marathon performance. Interaction terms between Age, and the environmental covariates allow for a detailed exploration of how the effects of these covariates on performance vary with age. The final model provides a comprehensive framework for understanding the complex relationships between environmental factors, age, and marathon performance.

$$\begin{aligned} \%CR = & \beta_0 + \beta_1 \cdot \text{Age} + \beta_2 \cdot \text{Age}^2 + \beta_3 \cdot \text{WBGT} + \beta_4 \cdot \text{Humidity} + \beta_5 \cdot \text{Solar Radiation} + \beta_6 \cdot \text{Wind} \\ & + \beta_7 \cdot (\text{Age:WBGT}) + \beta_8 \cdot (\text{Age:Humidity}) + \beta_9 \cdot (\text{Age:SR}) + \beta_{10} \cdot (\text{Age:Wind}) + \epsilon \end{aligned}$$

Table 3 highlights the significant covariates influencing marathon performance (%CR) among male participants. Age has a strong negative effect on performance, with an estimated decrease of 5.60 units in %CR per year, while its quadratic term ( $\text{Age}^2$ ) adds a non-linear effect of 0.08 units, indicating an accelerating decline in performance with increasing age. Among environmental factors, humidity has a positive effect on performance, with each unit increase in humidity improving %CR by 0.16 units. Conversely, solar radiation (SR) negatively affects performance, with each unit increase in SR leading to a 0.01 unit increase in %CR, reflecting worse performance.

The interaction terms further reveal how age modifies these environmental effects. For instance, the Age and WBGT interaction suggests that each additional year of age amplifies the negative impact of WBGT on performance by 0.01 units, while the Age and Humidity interaction indicates that the positive effect of humidity on performance diminishes by 0.004 units per year of age. Similarly, the Age and Solar Radiation interaction shows that the detrimental impact of SR becomes more pronounced with age, with a 0.003 unit increase in %CR for each year. Together, these results highlight a nuanced relationship where age and environmental conditions interact to influence marathon performance, with age exacerbating or mitigating the effects of key environmental factors.

Table 3: Summary of Significant Covariates (Men)

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	103.08	5.28	19.54	0.00
Age	-5.60	0.13	-44.04	0.00
I(Age <sup>2</sup> )	0.08	0.00	101.09	0.00
Humidity	0.16	0.05	3.21	0.00
SR	0.01	0.00	2.02	0.04
Age:WBGT	0.01	0.00	2.81	0.00
Age:Humidity	0.00	0.00	-4.87	0.00
Age:SR	0.00	0.00	-3.19	0.00

Table 4 presents the significant covariates for female participants. Age has a strong negative effect on performance, with an estimated decrease of 5.91 units in %CR per year, while its quadratic term ( $\text{Age}^2$ ) adds a non-linear effect of 0.08 units, confirming an accelerating decline in performance with age. Among environmental factors, WBGT significantly impacts performance, with each unit increase in WBGT associated with a 0.34 unit increase in %CR, reflecting worse performance. Unlike the male analysis, no other environmental variables or interactions appear significant in this model, suggesting that WBGT may be the primary environmental determinant of performance for women in this dataset.

The modeling results for men and women share key similarities but also highlight notable differences. Both models reveal a significant non-linear relationship between age and marathon performance, with the linear (Age) and quadratic (Age<sup>2</sup>) terms indicating a decline in performance with age at an increasing rate. Wet Bulb Globe Temperature (WBGT) is a significant factor for both genders, underscoring its broad impact on performance. However, men’s results show a more complex relationship with environmental variables, with significant effects for humidity, solar radiation (SR), and their interactions with age (Age:Humidity, Age:SR, Age:WBGT). In contrast, only WBGT is significant for women, and no age-environment interactions were observed. Additionally, women exhibit a slightly stronger decline in performance with age (-5.91) compared to men (-5.60). These findings suggest that while age-related trends are consistent across genders, men’s performance is more strongly influenced by environmental conditions, whereas women’s performance appears less sensitive to these factors, with WBGT being the dominant environmental influence.

Table 4: Summary of Significant Covariates (Women)

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	124.81	5.54	22.55	0.00
Age	-5.91	0.14	-41.12	0.00
I(Age <sup>2</sup> )	0.08	0.00	86.29	0.00
WBGT	0.34	0.17	2.03	0.04

## Discussion

Based on previous research on marathon performance, we expected performance to vary by age, gender, environmental conditions, or combinations of them. The exploratory analyses in this report suggest a “U-shaped” relationship between marathon performance and age, whereby performance initially rises, peaks at a certain age, and then declines. On average, women reach their highest performance at age 28 and men at age 30. Environmental conditions considered in the analysis included WBGT, humidity, solar radiation, and wind speed. Regardless of these environmental conditions, the 25-39 age group always outperformed the other age groups, and the oldest age group always had the highest %CR, meaning that they took the longest to complete the race. There were also some patterns between these four environmental conditions and the performance (%CR) of the oldest age group. For example, when the WBGT was increased from 0 to 10, the performance of both males and females in the oldest age group increased. However, no clear pattern was observed from the interaction plots, suggesting that weather conditions tended to affect younger participants.

There are several limitations in this study. First, the number of observations for the youngest and oldest age groups is relatively small compared to the middle-aged groups (Table 1). This may lead to a reduction in statistical power for the youngest and the oldest age groups. The findings for the youngest and oldest groups may not be as generalizable to the broader population in these age ranges. Second, this data contains little information on marathon performance in extreme weather conditions. This reduces the ability to predict how environmental factors, such as high WBGT (Wet Bulb Globe Temperature) or solar radiation, might affect marathon performance under such conditions. Without sufficient data on extreme weather conditions, the analysis may underestimate the true impact of these conditions on performance. Further, the statistical model used in this study does not capture the non-linear relationship between performance and age. More complex models and model selection processes may be needed to address this structure.

## Reference

- Ely, M. R., Cheuvront, S. N., Roberts, W. O., & Montain, S. J. (2007). Impact of weather on marathon-running performance. *Medicine and science in sports and exercise*, 39(3), 487-493.
- Kenney, W. L., & Munce, T. A. (2003). Invited review: aging and human temperature regulation. *Journal of applied physiology*, 95(6), 2598-2603.
- Besson, T., Macchi, R., Rossi, J., Morio, C. Y., Kunimasa, Y., Nicol, C., ... & Millet, G. Y. (2022). Sex differences in endurance running. *Sports medicine*, 52(6), 1235-1257.
- Yanovich, R., Ketko, I., & Charkoudian, N. (2020). Sex differences in human thermoregulation: relevance for 2020 and beyond. *Physiology*, 35(3), 177-184.
- Zavorsky, G. S., Tomko, K. A., & Smoliga, J. M. (2017). Declines in marathon performance: Sex differences in elite and recreational athletes. *PLoS ONE*, 12(2), e0172121. <https://doi.org/10.1371/journal.pone.0172121>

## Code Appendix

```
knitr::opts_chunk$set(echo = FALSE, warning = FALSE, message = FALSE)
# data manipulation
library(dplyr)
library(tidyr)

# summary tables
library(kableExtra)
library(gtsummary)
library(gt)

# interaction plots
library(ggplot2)
library(ggpubr)
library(ggExtra)
library(ggcorrplot)
library(lubridate)
library(gggridges)

# model
# library(lme4)
library(lmerTest)
library(broom.mixed)
# Load datasets
AQI <- read.csv("~/Desktop/PHP 2550/Data/aqi_values.csv")
Course <- read.csv("~/Desktop/PHP 2550/Data/course_record.csv")
Marathon <- read.csv("~/Desktop/PHP 2550/Data/marathon_dates.csv")
Project1<- read.csv("~/Desktop/PHP 2550/Data/project1.csv")

# Modify categorical variables
Project1 <- Project1 %>%
  rename(Race = Race..0.Boston..1.Chicago..2.NYC..3.TC..4.D.,
         Sex = Sex..0.F..1.M.)
Project1$Race <- as.factor(Project1$Race)
Project1$Sex <- as.factor(Project1$Sex)
Project1$Flag[Project1$Flag == ""] <- NA
Project1$Flag <- as.factor(Project1$Flag)
# Create categorical variable age_grp for age and modify humidity variable

# Distribution of humidity
ggplot(Project1, aes(x = X.rh)) +
  geom_histogram(fill = "skyblue", color="black") +
  labs(x = "Percent Relative Humidity", y = "Count", title = "Figure 1: Histogram of Percent Relative Humidity")
theme_minimal()

Project1 <- Project1 %>%
  mutate(age_grp = cut(Age..yr., breaks = c(0, 24, 39, 54, 69, Inf),
                      include.lowest = TRUE,
                      labels = c("< 25 yrs", "25-39 yrs", "40-54 yrs", "55-69 yrs", ">= 70 yrs"),
                      age_grp = factor(age_grp, levels = c("< 25 yrs", "25-39 yrs", "40-54 yrs", "55-69 yrs", ">= 70 yrs"),
                      X.rh = ifelse(X.rh <= 1, X.rh*100, X.rh))
```

```

# Summary Table 1 (N by gender and age group)
Project1 %>%
  select(Sex, age_grp) %>%
  mutate(Sex = ifelse(Sex==0, "Female", "Male")) %>%
  tbl_summary(by = Sex,
    label = list(age_grp ~ "Age Group"),
    statistic = list(
      all_continuous() ~ "{mean} ({sd})",
      all_categorical() ~ "{n}"),
    missing = "ifany",
    missing_text = "Missing") %>%
  modify_spanning_header(update = all_stat_cols() ~ "**Gender**") %>%
  modify_footnote(update = everything() ~ NA) %>%
  bold_labels() %>%
  as_kable_extra(booktabs = TRUE,
    caption = "Number of Participants by Age Group and Gender",
    longtable = TRUE, linesep = "") %>%
  kableExtra::kable_styling(font_size = 10,
    latex_options = c("repeat_header", "HOLD_position", "scale_down"))

# Summary Table 2
AQI_mod <- AQI %>%
  mutate(Race= case_when(marathon == "Boston" ~ "0",
    marathon == "Chicago" ~ "1",
    marathon == "NYC" ~ "2",
    marathon == "Twin Cities" ~ "3",
    marathon == "Grandmas" ~ "4",
  )) %>%
  filter(sample_duration != "1 HOUR") %>%
  group_by(Race, date_local) %>%
  summarize(ave_aqi = mean(aqi, na.rm = TRUE), .groups = "drop") %>%
  mutate(year = year(ymd(date_local))) %>%
  select(Race, ave_aqi, year)

Course_mod <- Course %>%
  mutate(Race= case_when(Race == "B" ~ "0",
    Race == "C" ~ "1",
    Race == "NY" ~ "2",
    Race == "TC" ~ "3",
    Race == "D" ~ "4",
  ),
  CR = period_to_seconds(hms(CR)))

# Merge Data
Project1_merged <- Project1 %>%
  left_join(Course_mod, by=c("Race", "Year")) %>%
  left_join(AQI_mod, by = c("Race", "Year" = "year"))

Project1_unique <- Project1_merged %>%
  distinct(Race, Year, .keep_all = TRUE)

Project1_tbl_w <- Project1_unique %>%
  select(Race, Flag, Td..C, Tw..C, X.rh, Tg..C, SR.W.m2, Wind, DP, CR, ave_aqi) %>%
  mutate(
    Race = case_when(Race == 0 ~ "Boston Marathon",

```

```

      Race == 1 ~ "Chicago Marathon",
      Race == 2 ~ "New York City Marathon",
      Race == 3 ~ "Twin Cities Marathon",
      Race == 4 ~ "Grandma's Marathon",
      TRUE ~ "Missing"),
  Flag = case_when(Flag == 'White' ~ "WBGT <10C",
    Flag == 'Green' ~ "WBGT 10-18C",
    Flag == 'Yellow' ~ "WBGT >18-23C",
    Flag == 'Red' ~ "WBGT >23-28C",
    TRUE ~ "Missing"),
  Flag = factor(Flag, levels = c("WBGT <10C", "WBGT 10-18C", "WBGT >18-23C", "WBGT >23-28C", "Missing"),
tbl_summary(by=Race,
  label = list(
    #Age..yr. ~ "Age",
    #X.CR ~ "Percent off current course record ",
    Td..C ~ "Dry bulb temperature",
    Tw..C ~ "Wet bulb temperature",
    X.rh ~ "Percent relative humidity",
    Tg..C ~ "Black globe temperature",
    SR.W.m2 ~ "Solar radiation in Watts",
    DP ~ "Dew Point",
    CR ~ "Course Record in seconds",
    ave_aqi ~ "Average Air Quality"
  ),
  statistic = list(all_continuous() ~ "{mean} ({sd})"),
  missing = "ifany",
  missing_text = "Missing") %>%
modify_spanning_header(update = all_stat_cols() ~ "***Race**") %>%
modify_footnote(update = all_stat_cols() ~ "Mean (SD) for continuous; n (%) for categorical") %>%
bold_labels()

Project1_tbl_w %>%
  as_kable_extra(booktabs = TRUE,
    caption = "Summary Statistics of Weather Condistions by Race",
    longtable = TRUE, linesep = "") %>%
kableExtra::kable_styling(font_size = 10,
  latex_options = c("repeat_header", "HOLD_position", "scale_down")) %>%
column_spec(1, width = "3cm") %>%
column_spec(2, width = "2cm") %>%
column_spec(3, width = "2cm") %>%
column_spec(4, width = "2cm") %>%
column_spec(5, width = "2cm") %>%
column_spec(6, width = "2cm") %>%
row_spec(0, bold = TRUE)
# Mean CR% by Age (male)
male_summary <- Project1 %>%
  filter(Sex == 1) %>%
  group_by(Age..yr.) %>%
  summarise(
    mean_CR = mean(X.CR, na.rm = TRUE),
    se_CR = sd(X.CR, na.rm = TRUE)
    # se_CR = sd(X.CR, na.rm = TRUE) / sqrt(n())
  )

```

```

# Mean CR% by Age (female)
female_summary <- Project1 %>%
  filter(Sex == 0) %>%
  group_by(Age..yr..) %>%
  summarise(
    mean_CR = mean(X.CR, na.rm = TRUE),
    se_CR = sd(X.CR, na.rm = TRUE)
    # se_CR = sd(X.CR, na.rm = TRUE) / sqrt(n())
  )

## 1.3 Interaction Plot (Age and %CR by gender)
# Plot (Male)
p_age_male <- ggplot(male_summary, aes(x = Age..yr., y = mean_CR)) +
  geom_point(color = "grey", size = 1) + # Add points for the mean
  geom_errorbar(aes(ymin = mean_CR - se_CR, ymax = mean_CR + se_CR), width = 1, color = "grey") + # Add error bars
  geom_smooth(se = FALSE, color = "black", size = 1, method = "loess", linetype = 2) + # Add smooth line
  geom_vline(xintercept = 70, linetype = "dashed", color = "red", size = 0.8) +
  annotate("text", x = 65, y = 300, label = "Age = 70", color = "red", angle = 90, size = 4, hjust = 0) +
  geom_point(aes(x = 30, y = 7.13), color = "blue", size = 2) +
  geom_hline(yintercept = 7.13, linetype = "dashed", color = "blue", size = 0.8) +
  annotate("text", x = 25, y = 60, label = "(30, 7.13)", color = "blue", angle = 0, size = 4, hjust = 0) +
  labs(title = "Men", x = "Age (yrs)", y = "Best Time (%CR)") +
  ylim(0, 400) +
  theme_minimal(base_size = 15) +
  theme(plot.title = element_text(hjust = 0.5, size=14),
        axis.title.x = element_text(size = 12),
        axis.title.y = element_text(size = 12))

# Plot (Female)
p_age_female <- ggplot(female_summary, aes(x = Age..yr., y = mean_CR)) +
  geom_point(color = "grey", size = 1) + # Add points for the mean
  geom_errorbar(aes(ymin = mean_CR - se_CR, ymax = mean_CR + se_CR), width = 1, color = "grey") + # Add error bars
  geom_smooth(se = FALSE, color = "black", size = 1, method = "loess", linetype = 2) + # Add smooth line
  geom_vline(xintercept = 70, linetype = "dashed", color = "red", size = 0.8) +
  annotate("text", x = 65, y = 300, label = "Age = 70", color = "red", angle = 90, size = 4, hjust = 0) +
  geom_point(aes(x = 28, y = 12.27), color = "blue", size = 2) +
  geom_hline(yintercept = 12.27, linetype = "dashed", color = "blue", size = 0.8) +
  annotate("text", x = 25, y = 60, label = "(28, 12.27)", color = "blue", angle = 0, size = 4, hjust = 0) +
  labs(title = "Women", x = "Age (yrs)", y = "Best Time (%CR)") +
  ylim(0, 400) +
  theme_minimal(base_size = 15) +
  theme(plot.title = element_text(hjust = 0.5, size=14),
        axis.title.x = element_text(size = 12),
        axis.title.y = element_text(size = 12))

# Merge
ggarrange(p_age_male+ rremove("ylab") + rremove("xlab"),
          p_age_female+ rremove("ylab") + rremove("xlab"),
          ncol = 2, nrow = 1,
          labels = c("A", "B"),
          font.label = list(size = 12),
          common.legend = TRUE, align = "hv") %>%
  annotate_figure(
    top = text_grob("Figure 2: Interaction between %CR and Age by Gender",

```

```

      face = "bold", size = 14),
    left = text_grob("Best Time (%CR)", rot = 90),
    bottom = text_grob("Age (yrs)" )
## 1.2 Summary Statistics of %CR by Age Group and Gender
male_summary_grp <- Project1 %>%
  filter(Sex == 1) %>%
  group_by(age_grp) %>%
  summarise(
    mean_CR = mean(X.CR, na.rm = TRUE), # Compute the mean
    lower = quantile(X.CR, 0.25, na.rm = TRUE), # Compute the lower quantile (25%)
    upper = quantile(X.CR, 0.75, na.rm = TRUE) # Compute the upper quantile (75%)
  ) %>%
  mutate(
    Male = paste0(round(mean_CR, 2), " (", round(lower, 2), ", ", round(upper, 2), ")") # Combine into
  ) %>%
  select(age_grp, Male)

female_summary_grp <- Project1 %>%
  filter(Sex == 0) %>%
  group_by(age_grp) %>%
  summarise(
    mean_CR = mean(X.CR, na.rm = TRUE), # Compute the mean
    lower = quantile(X.CR, 0.25, na.rm = TRUE), # Compute the lower quantile (25%)
    upper = quantile(X.CR, 0.75, na.rm = TRUE) # Compute the upper quantile (75%)
  ) %>%
  mutate(
    Female = paste0(round(mean_CR, 2), " (", round(lower, 2), ", ", round(upper, 2), ")") # Combine into
  ) %>%
  select(age_grp, Female)

# Combine two summary tables
summary_sex_grp <- cbind(male_summary_grp, female_summary_grp)[, -3]
empty_row <- tibble(age_grp = "CR% (by Age Group)", Male = "", Female = "")
summary_sex_grp <- bind_rows(empty_row, summary_sex_grp)

summary_sex_grp %>%
  gt() %>%
  cols_label(
    age_grp = "Characteristic",
    Male = "Men, N=6,112",
    Female = "Women, N=5,452") %>%
  tab_header(
    title = md("Table 3: Summary Statistics of %CR by Gender, Mean (Q1, Q3)") %>%
  tab_style(
    style = cell_text(size = px(6), weight = "bold"), # Bold the column labels
    locations = cells_column_labels(columns = c(age_grp, Male, Female))) %>%
  tab_style(
    style = cell_text(size = px(6), weight = "bold"), # Bold the "CR% (by age group)" row
    locations = cells_body(columns = everything(), rows = age_grp == "CR% by Age Group")) %>%
  tab_style(
    style = cell_text(size = px(8), weight = "bold"), # Adjust the header size
    locations = cells_title(groups = "title") # Apply to the title only
  ) %>%

```



```

tab_options(
  table.width = pct(100),
  table.align = "center"
)
Project1_name <- Project1_merged %>%
  rename(
    #"% CR" = X.CR,
    "Dry bulb Temp" = Td..C,
    "Wet bulb Temp" = Tw..C,
    "Percent relative humidity" = X.rh,
    "Globe Temp" = Tg..C,
    "Solar Radiation" = SR.W.m2,
    "Average AQI" = ave_aqi,
    "Dew Point" = DP
  )
r <- cor(Project1_name[, c(6:14,17)], use="complete.obs")

variable_order <- c("Dry bulb Temp", "Wet bulb Temp",
  "Globe Temp", "Solar Radiation", "Wind",
  "Percent relative humidity", "Average AQI", "Dew Point", "WBGT"
  #, "% CR"
)

# Reorder the correlation matrix based on the new variable order
r_reordered <- r[variable_order, variable_order]

ggcorrplot(r_reordered,
  hc.order = TRUE,
  lab_size = 2.5,
  type = "lower",
  lab = TRUE) +
  ggtitle("Figure 3: Correlation Matrix of Environmental Variables") +
  theme(plot.title = element_text(hjust = 0.5, size=10),
    axis.text.x = element_text(size = 8),
    axis.text.y = element_text(size = 8))

## 2.1 WBGT (Temperature)
women_summary_WBGT <- Project1 %>%
  filter(Sex == 0) %>%
  group_by(WBGT, age_grp) %>%
  summarise(
    mean_CR = mean(X.CR, na.rm = TRUE)
    # ,
    # se_CR = sd(X.CR, na.rm = TRUE)
  ) %>%
  mutate(Sex = "Women")

men_summary_WBGT <- Project1 %>%
  filter(Sex == 1) %>%
  group_by(WBGT, age_grp) %>%
  summarise(
    mean_CR = mean(X.CR, na.rm = TRUE)
    # ,
    # se_CR = sd(X.CR, na.rm = TRUE)
  )

```

```

) %>%
mutate(Sex = "Men")

# Combine data for both sexes
combined_summary_WBGT <- bind_rows(women_summary_WBGT, men_summary_WBGT)

# Create a single plot
plot_WBGT <- ggplot(combined_summary_WBGT, aes(x = WBGT, y = mean_CR, color = age_grp, linetype = Sex)) +
  geom_smooth(se = FALSE, size = 0.7, method = "loess") + # Add smooth line without confidence intervals
  labs(
    #title = "Figure 4: Interaction between %CR and WBGT with Age by Gender",
    x = "WBGT",
    y = " ",
    color = "Age Group",
    linetype = "Gender"
  ) +
  ylim(0, 200) + # Set y-axis limit
  theme_minimal(base_size = 15) + # Use minimal theme with larger font
  theme(
    plot.title = element_text(hjust = 0.5, size = 14),
    axis.title.x = element_text(size = 12),
    axis.title.y = element_text(size = 12),
    legend.text = element_text(size = 5),
    legend.title = element_text(size = 6)
  )
)

women_summary_hum <- Project1 %>%
  filter(Sex == 0) %>%
  group_by(X.rh, age_grp) %>%
  summarise(
    mean_CR = mean(X.CR, na.rm = TRUE)
    # ,
    # se_rh = sd(X.rh, na.rm = TRUE)
  ) %>%
  mutate(Sex = "Women")

men_summary_hum <- Project1 %>%
  filter(Sex == 1) %>%
  group_by(X.rh, age_grp) %>%
  summarise(
    mean_CR = mean(X.CR, na.rm = TRUE)
    # ,
    # se_rh = sd(X.rh, na.rm = TRUE)
  ) %>%
  mutate(Sex = "Men")

# Combine data for both sexes
combined_summary_hum <- bind_rows(women_summary_hum, men_summary_hum)

# Create a single plot
plot_hum <- ggplot(combined_summary_hum, aes(x = X.rh, y = mean_CR, color = age_grp, linetype = Sex)) +
  geom_smooth(se = FALSE, size = 0.7, method = "loess") + # Add smooth line without confidence intervals
  labs(
    x = "Percent relative humidity",

```

```

    y = " ",
    color = "Age Group",
    linetype = "Gender"
  ) +
  ylim(0, 175) + # Set y-axis limit
  theme_minimal(base_size = 15) + # Use minimal theme with larger font
  theme(
    plot.title = element_text(hjust = 0.5, size = 14),
    axis.title.x = element_text(size = 12),
    axis.title.y = element_text(size = 12),
    legend.text = element_text(size = 5),
    legend.title = element_text(size = 6)
  )
women_summary_SR <- Project1 %>%
  filter(Sex == 0) %>%
  group_by(SR.W.m2, age_grp) %>%
  summarise(
    mean_CR = mean(X.CR, na.rm = TRUE),
    se_CR = sd(X.CR, na.rm = TRUE)
  ) %>%
  mutate(Sex = "Women")

men_summary_SR <- Project1 %>%
  filter(Sex == 1) %>%
  group_by(SR.W.m2, age_grp) %>%
  summarise(
    mean_CR = mean(X.CR, na.rm = TRUE),
    se_CR = sd(X.CR, na.rm = TRUE)
  ) %>%
  mutate(Sex = "Men")

# Combine data for both sexes
combined_summary_SR <- bind_rows(women_summary_SR, men_summary_SR)

# Create a single plot
plot_SR <- ggplot(combined_summary_SR, aes(x = SR.W.m2, y = mean_CR, color = age_grp, linetype = Sex)) +
  geom_smooth(se = FALSE, size = 0.7, method = "loess") + # Add smooth line without confidence intervals
  labs(
    x = "Solar Radiatio",
    y = " ",
    color = "Age Group",
    linetype = "Gender"
  ) +
  ylim(0, 175) + # Set y-axis limit
  theme_minimal(base_size = 15) + # Use minimal theme with larger font
  theme(
    plot.title = element_text(hjust = 0.5, size = 14),
    axis.title.x = element_text(size = 12),
    axis.title.y = element_text(size = 12),
    legend.text = element_text(size = 5),
    legend.title = element_text(size = 6)
  )
women_summary_wind <- Project1 %>%

```

```

filter(Sex == 0) %>%
group_by(Wind, age_grp) %>%
summarise(
  mean_CR = mean(X.CR, na.rm = TRUE),
  se_CR = sd(X.CR, na.rm = TRUE)
) %>%
mutate(Sex = "Women")

men_summary_wind <- Project1 %>%
filter(Sex == 1) %>%
group_by(Wind, age_grp) %>%
summarise(
  mean_CR = mean(X.CR, na.rm = TRUE),
  se_CR = sd(X.CR, na.rm = TRUE)
) %>%
mutate(Sex = "Men")

# Combine data for both sexes
combined_summary_wind <- bind_rows(women_summary_wind, men_summary_wind)

# Create a single plot
plot_wind <- ggplot(combined_summary_wind, aes(x = Wind, y = mean_CR, color = age_grp, linetype = Sex))
  geom_smooth(se = FALSE, size = 0.7, method = "loess") + # Add smooth line without confidence intervals
  labs(
    x = "Wind",
    y = " ",
    color = "Age Group",
    linetype = "Gender"
  ) +
  ylim(0, 175) + # Set y-axis limit
  theme_minimal(base_size = 15) + # Use minimal theme with larger font
  theme(
    plot.title = element_text(hjust = 0.5, size = 14),
    axis.title.x = element_text(size = 12),
    axis.title.y = element_text(size = 12),
    legend.text = element_text(size = 5),
    legend.title = element_text(size = 6)
  )

# Combine the plots into a single figure
combined_plot <- ggarrange(
  plot_WBGT,
  plot_hum,
  plot_SR,
  plot_wind,
  common.legend = TRUE,
  legend = "bottom",
  labels = c("A", "B", "C", "D"),
  label.font = list(size = 6),
  ncol = 2,
  nrow = 2
)

```

```

# Annotate the combined plot without adding extra layers
annotate_figure(
  combined_plot$`1`,
  top = text_grob("Figure 4: Environmental Factors and Marathon Performance", face = "bold", size = 10),
  left = text_grob("%CR", rot = 90, size = 10)
)
Project1_merged_mod <- Project1_merged %>%
  rename(
    Age = `Age.yr`,
    Humidity = `X.rh`,
    SR = `SR.W.m2`,
    AQI = `ave_aqi`
  )
# model
M1_men <- lm(X.CR ~ Age + I(Age^2) +
             (WBGT + Humidity + SR + Wind + AQI) +
             Age*(WBGT + Humidity + SR + Wind + AQI),
             data = Project1_merged_mod[Project1_merged_mod$Sex == 1,])

# Create coefficient estimate table
## Men
model_summary_men <- summary(M1_men)
coefficients_data_men<- as.data.frame(model_summary_men$coefficients)
colnames(coefficients_data_men) <- c("Estimate", "Std. Error", "t value", "Pr(>|t|)")
coefficients_data_men$Coefficients <- rownames(coefficients_data_men)
coefficients_data_men <- coefficients_data_men[, c("Coefficients", "Estimate", "Std. Error", "t value",
  filter(`Pr(>|t|)`<0.05) %>%
  mutate(across(where(is.numeric), ~ round(., 2))))

coefficients_data_men[, -1] %>%
  kbl(booktabs = TRUE, caption = "Summary of Significant Covariates (Men)",
      #col.names = c("Group", "Term", "SD"),
      longtable = TRUE, linesep = "") %>%
  kable_styling(font_size = 10,
                latex_options = c("repeat_header", "HOLD_position", "scale_down"))
M1_women <- lm(X.CR ~ Age + I(Age^2) +
              (WBGT + Humidity + SR + Wind + AQI) +
              Age*(WBGT + Humidity + SR + Wind + AQI),
              data = Project1_merged_mod[Project1_merged_mod$Sex == 0,])

## Women
model_summary_women <- summary(M1_women)
coefficients_data_women<- as.data.frame(model_summary_women$coefficients)
colnames(coefficients_data_women) <- c("Estimate", "Std. Error", "t value", "Pr(>|t|)")
coefficients_data_women$Coefficients <- rownames(coefficients_data_women)
coefficients_data_women <- coefficients_data_women[, c("Coefficients", "Estimate", "Std. Error", "t val
  filter(`Pr(>|t|)`<0.05)%>%
  mutate(across(where(is.numeric), ~ round(., 2))))

coefficients_data_women[, -1] %>%
  kbl(booktabs = TRUE, caption = "Summary of Significant Covariates (Women)",
      #col.names = c("Group", "Term", "SD"),
      longtable = TRUE, linesep = "") %>%
  kable_styling(font_size = 10,

```

```
latex_options = c("repeat_header", "HOLD_position", "scale_down"))
```