Examining the Impact of Environmental and Demographic Factors on Marathon Performance: An Exploratory Data Analysis

Michael Stewart

2024-10-06

Introduction

This project, conducted in collaboration with Dr. Brett Romano Ely and Dr. Matthew Ely from Providence College, examines the effects of environmental conditions on marathon performance across age and gender. The dataset includes performances from five major marathons (Boston, New York, Chicago, Twin Cities, Grandma's) with detailed environmental data such as temperature, humidity, WBGT, and air quality, for 11,564 athletes aged 14 to 85.

Through exploratory data analysis, we aimed to investigate three main objectives: first, to assess the effects of aging on marathon performance in men and women; second, to explore how environmental conditions such as temperature and air quality affect performance; and third, to identify which specific environmental factors, including WBGT and temperature, have the greatest impact on performance. Our hypotheses suggest that older athletes and women may experience more significant declines in performance under adverse conditions. Statistical analyses, including ANOVA and correlation analysis, were employed to examine these relationships and provide insight into the interaction between age, gender, and environmental factors in endurance performance.

Methods

Missing Data Handling

Missing weather data were addressed through a complete case analysis, removing rows with missing values for key environmental variables, such as wet bulb temperature, dry bulb temperature, relative humidity, and air quality (ozone PPM). The proportion of missing data was minimal, accounting for only 4.25% of total observations, as shown in **Table 0**. For instance, in 2011, the Chicago marathon had 126 missing observations (1.09%), New York had 131 (1.13%), and Twin Cities had 118 (1.02%). Duluth (Grandma's) marathon in 2012 had 116 missing values (1.00%). Given the relatively small percentage of missing data, this approach ensured the integrity of the data set while avoiding unnecessary complexity in imputation. Dropping these missing values yielded 11,073 participants left to be analyzed.

Table 0: Summary of Missing Weather Data by Year and Race
Missing Data Relative to Total Observations

Race	Missing Data Count	Missing Data (%)
2011		
Chicago	126	1.09
New York	131	1.13
Twin Cities	118	1.02
2012		
Duluth (Grandma's)	116	1.00
Total		
	491	4.25

Data Preparation

The dataset was comprised of marathon performance data from five major races, along with corresponding environmental conditions, including air quality measures. Marathon performance data were merged with course records, race dates, and air quality data. The primary air quality variable, ozone concentration, was measured in parts per million (PPM) and calculated as an 8-hour average for each race-year. We chose PPM over micrograms per cubic meter because PPM is the standard unit for measuring gas concentrations like ozone, whereas micrograms per cubic meter is typically used for particulate matter such as PM2.5. The 8-hour average was selected to better reflect longer-term exposure during marathon events, as opposed to shorter-term (e.g., 1-hour) measurements, which may not capture the full extent of air quality conditions experienced by runners throughout the race. This approach provided a standardized metric for comparing air quality across different races and years.

To ensure comparability across different analyses, marathon times were treated in two distinct ways depending on the context. When analyses did not group or facet by race, marathon times were represented in minutes, allowing for more straightforward and interpretable results. However, when grouping or comparing across races, percentcourse recordwas used (percent CR) to account for differences in race conditions and course difficulty. Percent CR standardizes a runner's time as a percentage of the course record for each marathon. A value of 0% means the runner matched the course record, while positive values indicate slower times, and negative values represent faster times. This allows performance to be compared across different races and conditions.

Exploratory Data Analysis

Exploratory data analysis (EDA) focused on the relationship between marathon performance and the variables of interest: age, gender, and environmental conditions. Marathon times were analyzed using summary statistics and visualizations, such as boxplots, to explore trends across different age groups and genders. This analysis provided insights into how performance varied with age, with comparisons made across men and women.

Statistical Analysis

To assess the effects of environmental variables and demographic factors on marathon performance, a series of ANOVA models were employed. These models examined the influence of gender, age, and race-day flag conditions on performance times. Post-hoc Tukey tests were performed to identify significant pairwise differences across these groups. Interaction effects were also investigated between age and gender, as well as between age and flag conditions.

Results

Aim 1: Effects of Increasing Age on Marathon Performance in Men and Women

The analysis of marathon performance across age groups and genders reveals clear trends regarding how marathon times are affected by aging. As shown in **Figure 1.1**, marathon times are fastest among runners in the 26-35 age range across all five major marathons. Both men and women exhibit slower times as they age, with performance declining progressively in the older age groups. The gender differences are evident, with men generally running faster than women at all ages. Notably, the gap between male and female performance becomes more pronounced in the older age categories, particularly among runners aged 56-64 and 65+, where women tend to take significantly longer to complete the marathon compared to men. Variability in times also increases as participants age for both sexes. It's important to note that New York and Boston did not record any times for participants in the <18 age group, which may influence the comparisons for this category across marathons.

Figure 1.1: Marathon Times by Age and Gender

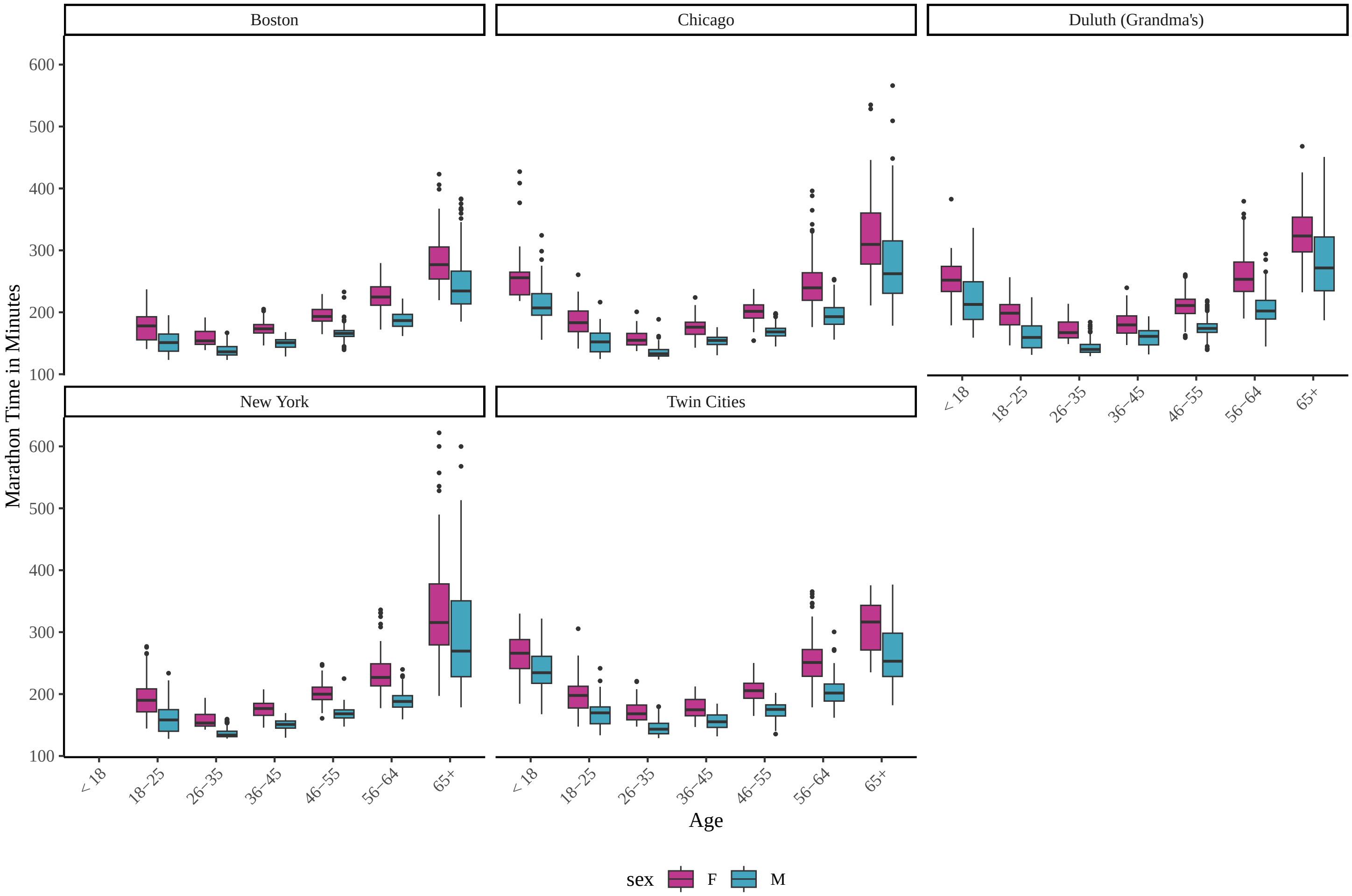


Table 1.1, summarizing marathon performance by age and sex, further illustrates these trends. For instance, among 26-35 year olds, male marathoners averaged 139.4 minutes (SD = 10.1), while female marathoners averaged 162.8 minutes (SD = 15.4). This pattern persists across all age groups, and by the 65+ category, men were completing marathons in an average of 276 minutes (SD = 67.6) compared to women's 318.1 minutes (SD = 63.3). This nearly 42-minute difference shows the increasing performance gap as athletes age.

Table 1.1: Marathon Performance Summary by Age Group and Sex

Age Group	Male (n, %)	Female (n, %)	Male Marathon Time[1]	Female Marathon Time[1]	n
< 18	158 (58.1%)	114 (41.9%)	226 (38.8)	259.1 (39.5)	272
18-25	732 (50%)	732 (50%)	158.4 (20.4)	188.8 (26.4)	1,464
26-35	920 (50%)	920 (50%)	139.4 (10.1)	162.8 (15.4)	1,840
36-45	920 (50%)	920 (50%)	153.4 (11.3)	176 (15.4)	1,840
46-55	920 (50%)	920 (50%)	170 (12)	202.4 (17.2)	1,840
56-64	919 (50.5%)	901 (49.5%)	195.5 (19.5)	242.5 (34.3)	1,820
65+	1286 (64.4%)	711 (35.6%)	276 (67.6)	318.1 (63.3)	1,997

[1]Marathon time is presented as mean (SD) in minutes.

The ANOVA results (**Table 1.2**) show that both age and sex have statistically significant effects on marathon times ($p < 0.001$). Gender alone accounts for a large portion of the variance (F-value = 1306.618, $p < 0.001$), and age also plays a significant role (F-value = 4112.184, $p < 0.001$). Also, there is a significant interaction between gender and age (F-value = 33.917, $p < 0.001$), suggesting that the rate of decline with age differs between men and women. The interaction between age and sex is also significant ($p < 0.001$), indicating that the rate of performance decline with age differs between men and women. This interaction is further explored in the Tukey post-hoc analysis (**Table 1.3**), which shows that men outperform women across all age groups, with the largest differences observed in the oldest categories. For example, the difference between males and females in the 18-25 age group is -22.853 minutes ($p < 0.001$), while the difference is even greater in the 65+ group, where females take, on average, 52.56 minutes longer to complete the marathon compared to males. Also, the fact that the difference between the 26-35 and 18-25 groups is smaller than those involving older groups (−22.534 minutes, p < 0.0001) indicates that peak marathon performance falls within these ranges.

Table 1.2: ANOVA Results for Marathon Times by Flag Condition, Gender, and Age Bin

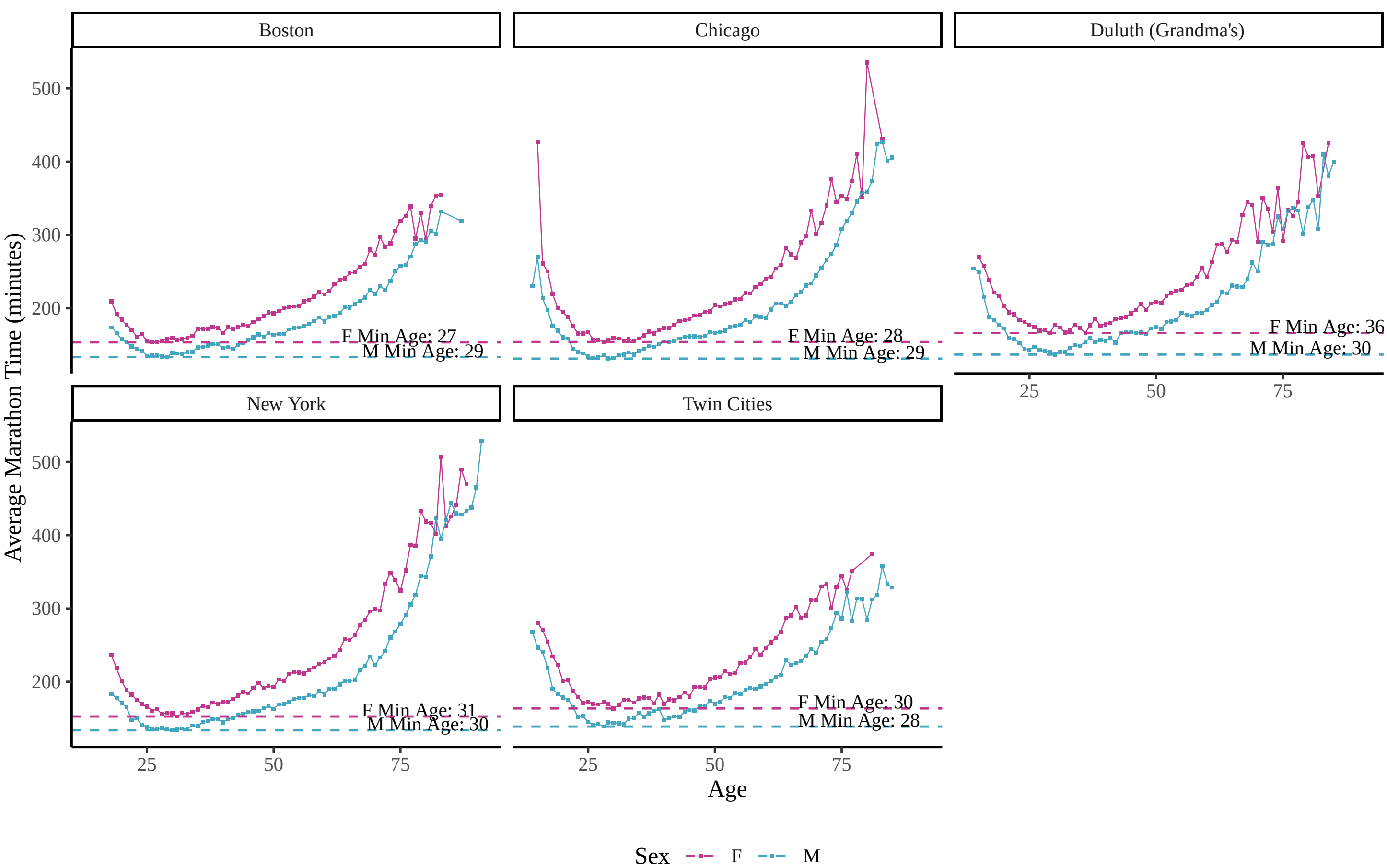
Source	Df	Sum Sq	Mean Sq	F-value	Pr(>F)
sex	1	1,441,001	1,441,001	1,306.618	0.000
age_bin	6	27,210,677	4,535,113	4,112.184	0.000
flag	3	190,009	63,336	57.430	0.000
sex:age_bin	6	224,430	37,405	33.917	0.000
sex:flag	3	14,399	4,800	4.352	0.005
age_bin:flag	18	37,468	2,082	1.887	0.013
sex:age_bin:flag	18	13,137	730	0.662	0.852
Residuals	11017	12,150,075	1,103	NA	NA

Table 1.3: Tukey Post-Hoc Test Results for Sex
95% Family-Wise Confidence Level

Difference	Lower Bound	Upper Bound	Adjusted p-value	Sex Comparison
−22.853	−24.093	−21.614	0.000	M-F

Further analysis, as visualized in **Figure 1.2**, shows that peak performance age varies between men and women across different races. Men tend to reach their peak performance around 28-30 years old, and women tend to reach there peak around 28 to 36 years. One particularly interesting finding is that in Duluth, women's best performance age extends to 36 years. This six-year gap compared to men, who generally peak around 29, raises questions about whether the summer conditions in Duluth might contribute to this outlier. This suggests that gender differences in thermoregulation may play a role in women reaching their peak later in Duluth races. Overall, the 26-35 age bin is the period of fastest marathon times for both men and women.

Figure 1.2: Average Marathon Performance by Age and Race



In summary, these results support the hypothesis that aging impacts marathon performance differently for men and women. Both men and women experience declines in performance as they age. These findings offer valuable insights into the interplay between age and sex in long-distance running events.

Table 1.3: Tukey Post-Hoc Test Results for Age Bins
95% Family-Wise Confidence Level

Difference	Lower Bound	Upper Bound	Adjusted p-value	Age Bin Comparison
−68.137	−74.603	−61.671	0.000	18-25-< 18
−90.671	−97.032	−84.309	0.000	26-35-< 18
−77.049	−83.411	−70.688	0.000	36-45-< 18
−55.501	−61.862	−49.139	0.000	46-55-< 18
−22.870	−29.236	−16.504	0.000	56-64-< 18
52.560	46.231	58.889	0.000	65+-< 18
−22.534	−25.963	−19.104	0.000	26-35-18-25
−8.912	−12.342	−5.483	0.000	36-45-18-25
12.636	9.207	16.066	0.000	46-55-18-25
45.267	41.829	48.705	0.000	56-64-18-25
120.697	117.328	124.067	0.000	65+-18-25
13.621	10.393	16.850	0.000	36-45-26-35
35.170	31.941	38.399	0.000	46-55-26-35
67.801	64.564	71.039	0.000	56-64-26-35
143.231	140.066	146.395	0.000	65+-26-35
21.549	18.320	24.777	0.000	46-55-36-45
54.180	50.942	57.417	0.000	56-64-36-45
129.609	126.445	132.774	0.000	65+-36-45
32.631	29.393	35.868	0.000	56-64-46-55
108.061	104.896	111.225	0.000	65+-46-55
75.430	72.256	78.603	0.000	65+-56-64

Aim 2: Effects of Environmental Conditions on Marathon Performance in Men and Women

For Aim 2, the goal was to determine how environmental factors influence marathon performance and whether these effects differ by age and gender. We determined 2 kinds of environmental conditions: weather and air quality. We first looked at the weather parameters.

Relationship between Weather Variables and Marathon Performance

Table 2.1 displays some summary statistics for all related weather variables.

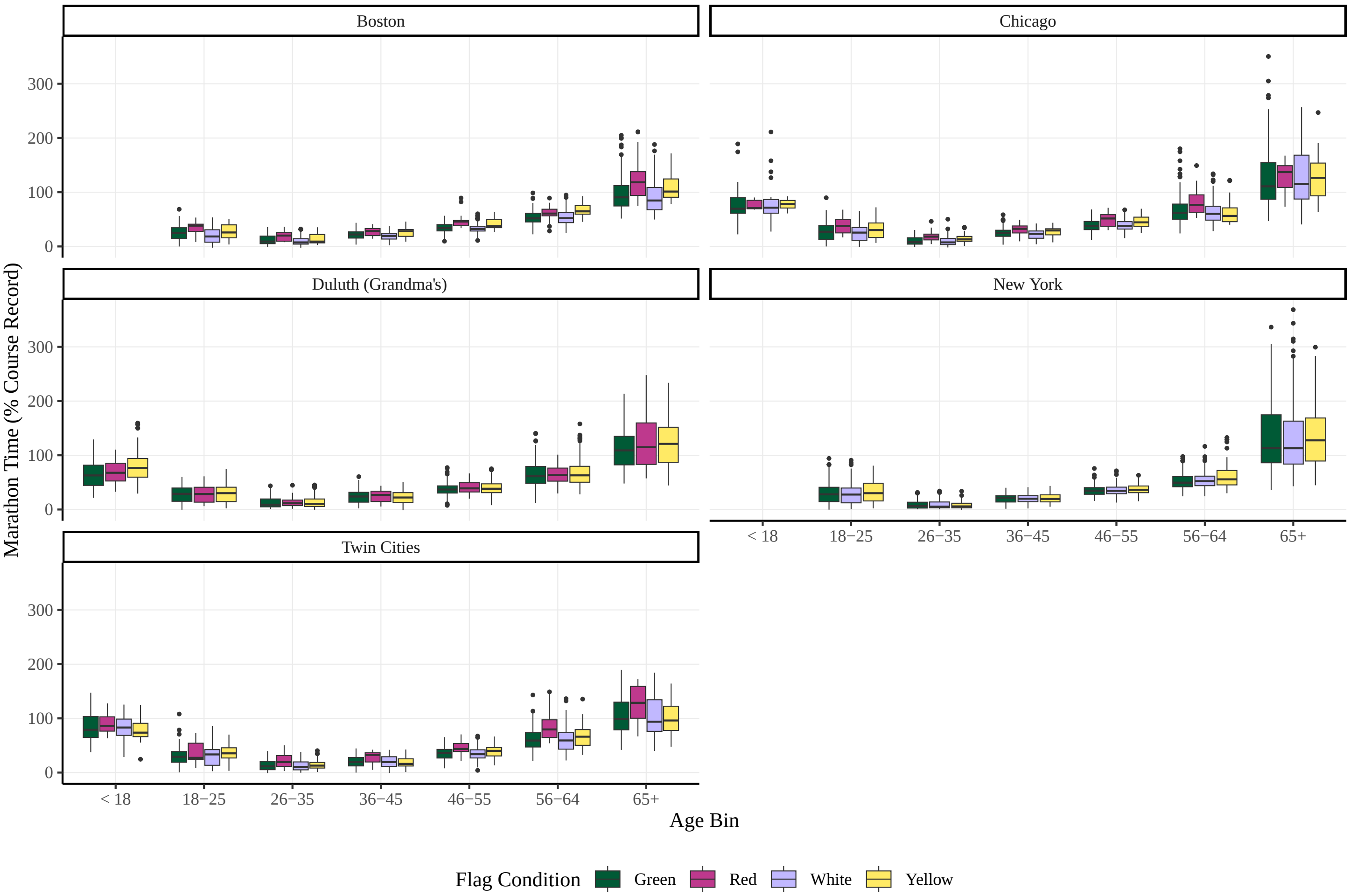
Table 2.1: Weather Summary by Race
Median (IQR) of Weather Variables

Race	WBT (°C)[1]	DBT (°C)[1]	RH (%)[1]	SR (W/m²)[1]	Wind (m/s)[1]	WBGT (°C)[1]
Boston	7.2 (5.4,8.2)	9 (8.3,13.8)	39.2 (0.6,58.2)	708.1 (574,800.3)	12 (8.3,16)	10.3 (8.7,12.7)
Chicago	9.4 (2.5,12.9)	14.5 (7,15.7)	59.5 (51.3,65.7)	479 (436.6,535.5)	8.4 (5.3,10.3)	13.6 (6.7,16.4)
Duluth (Grandma's)	14.3 (13.7,16.9)	18.1 (17,22)	60.3 (0.9,80.3)	731.3 (520,833.2)	9 (7,11.2)	18.1 (17.1,21.8)
New York	7.6 (2.9,11.5)	12 (7.4,15.1)	1 (0.4,55.3)	416.7 (309.5,546.2)	11 (9,14)	10.9 (6.7,14.1)
Twin Cities	8.5 (7.3,11.1)	11.3 (9,15.7)	55.5 (0.7,76.3)	481.3 (348,545.3)	9.3 (6.2,10)	12.5 (9,14.4)

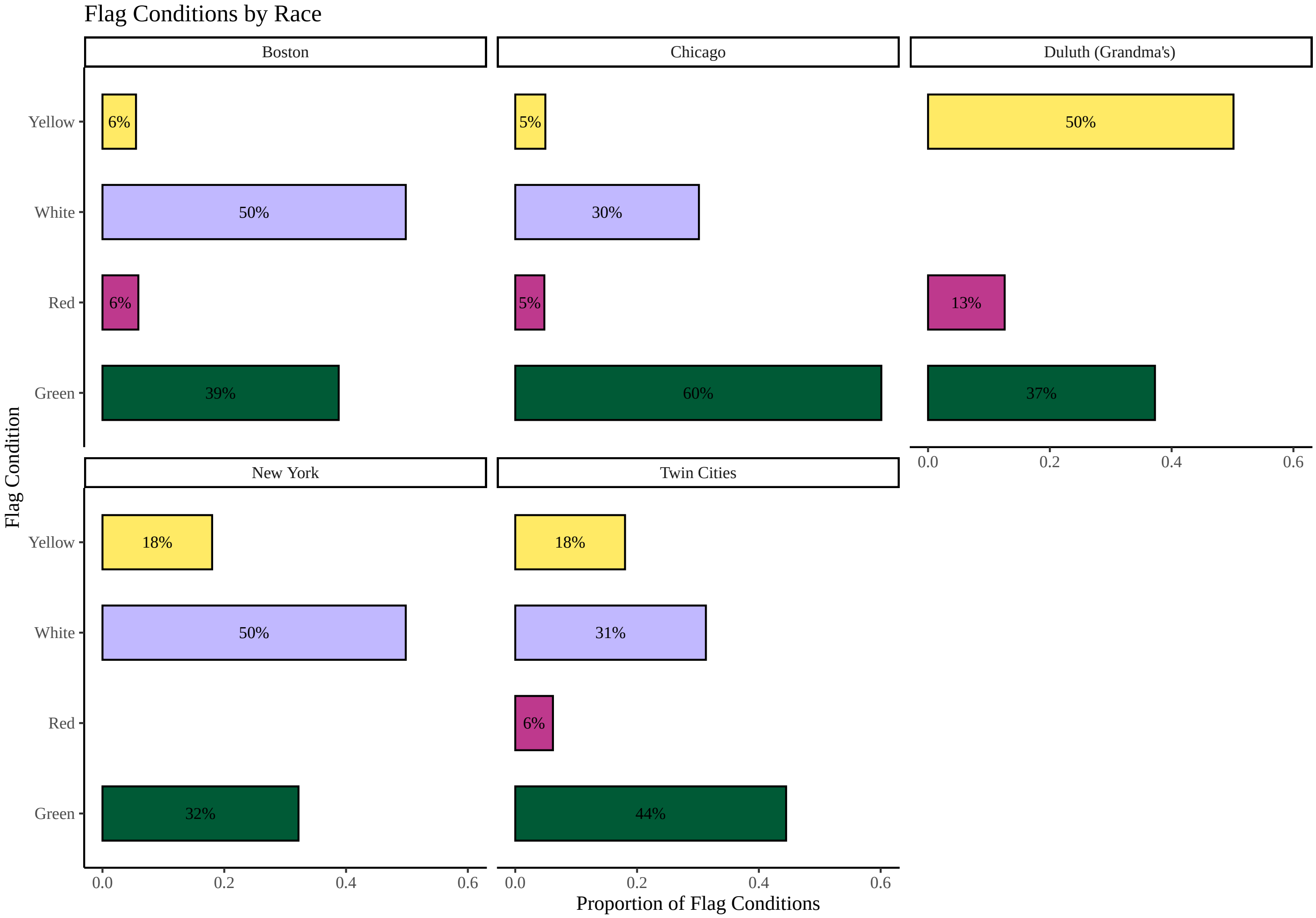
[1]WBT: Wet Bulb Temp, DBT: Dry Bulb Temp, RH: Relative Humidity, SR: Solar Radiation, WBGT: Wet Bulb Globe Temp

From this table, we can see some differences in weather conditions across different races. For example, Duluth (Grandma's), held during the summer, consistently experiences the highest median wet bulb temperature (14.3°C) and dry bulb temperature (18.1°C), along with the highest wet bulb globe temperature (WBGT) at 18.1°C. This race also has the highest relative humidity (60.3%), which, combined with high solar radiation and temperatures, suggests more challenging conditions for runners compared to other races. By contrast, New York and Boston races tend to have windier conditions, with median wind speeds of 11 and 12 m/s, respectively, compared to other races. These factors highlight Duluth's more intense environmental conditions, which may contribute to longer marathon times due to heat and humidity. We further explored the relationship between weather and marathon performance in **Figure 2.1**, where we incorporated flag conditions as a categorical representation of WBGT. Flag conditions serve as indicators of temperature-related risk, with white flags representing the coldest conditions, followed by green for moderate conditions, yellow for warmer, and red for the hottest and most hazardous conditions. The differences in weather conditions across races, particularly Duluth's harsher environment, may impact runners' performance and influence the distribution of flag conditions.

Figure 2.1: Marathon Performance by Age Bin and Flag Condition



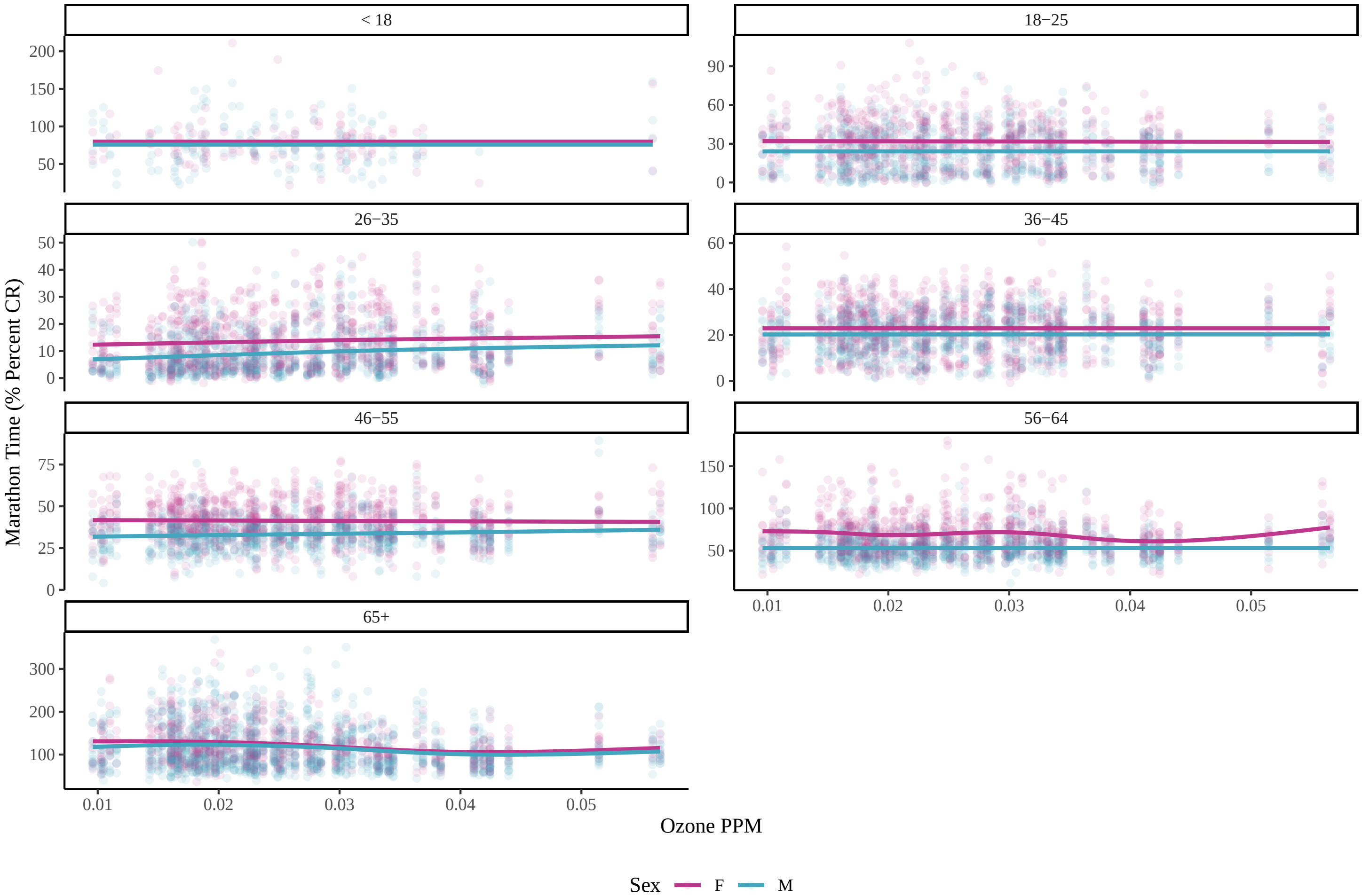
These results show that marathon times increase with age across all races, especially in the older age groups (65+). The red flags, indicating the hottest conditions, are generally associated with the highest marathon times, particularly in the 65+ age group, where there is considerable variation. The white and green flags, which represent cooler conditions, tend to have the fastest times across most age bins, highlighting the impact of favorable weather conditions on performance. We observe that New York does not have any races with red flags from the given years in the data, and Duluth does not have any white flag races. The latter is true, most likely because it takes place in the summer, where white flag conditions are extremely rare in that location. Also, we can see missingness in the < 18 age group in New York and Boston, where one could infer that they are missing because one must be 18+ to qualify for those marathons, or perhaps there is some other reason we don;t observe people from that age range in our data. **Figure 2.2** further illustrates the differences in flag distributions across races.



Relationship Between Air Quality and Marathon Performance

The second type of environmental factor that we explored was air quality, measured by ozone in parts per million (PPM). In **Figure 2.3**, we visualized the marathon times in percent course record across different PPM levels that were observed in the data, stratified by age range. For all age groups, we mostly saw a flat regression line fit to the data, indicating a weak linear relationship. There were minor fluctuations in older groups but they were still relatively flat and uninformative. From this visual, we assume that air quality most likely does not have a significant effect on marathon times for neither men nor women. Because of this, we will focus on discovering which weather parameters impact marathon performance the most in aim 3.

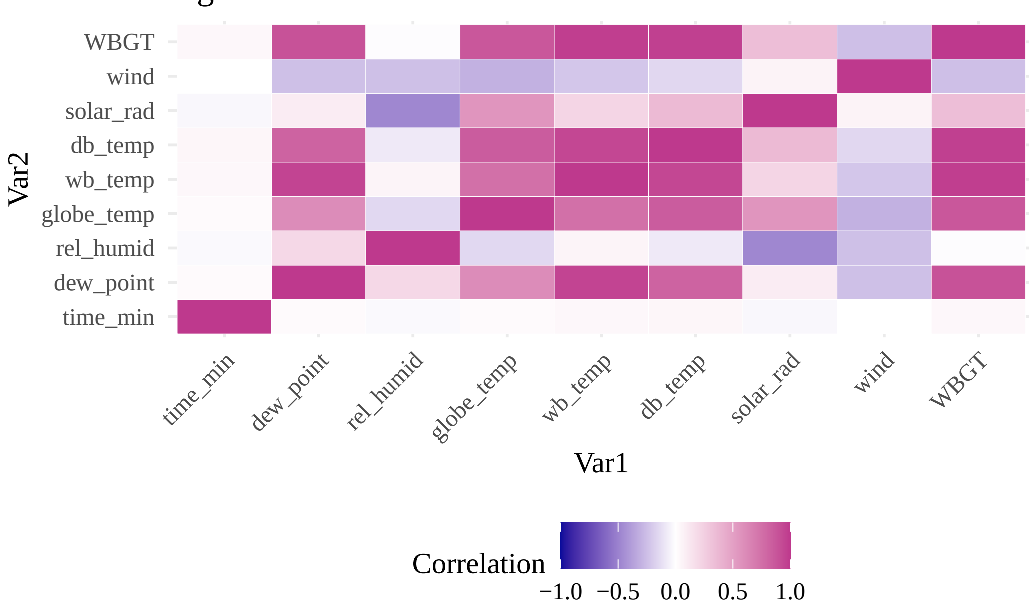
Figure 2.3: Air Quality (Ozone PPM) vs Marathon Time by Sex & Age Group



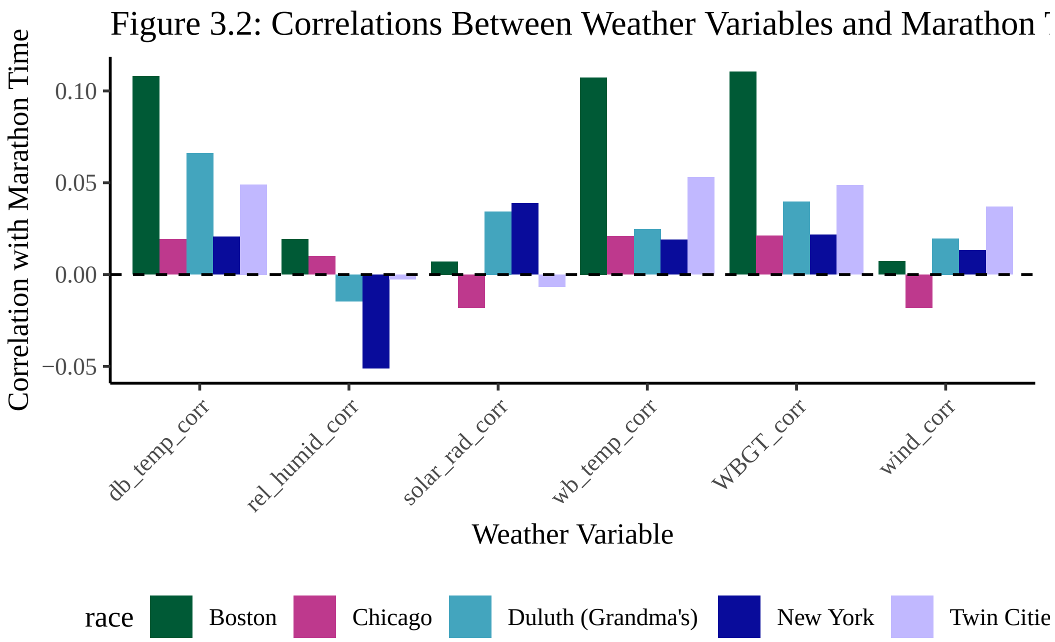
Aim 3

For Aim 3, the analysis shifted from comparing the effects of environmental conditions across demographic groups to identifying which specific weather variables had the most significant impact on overall marathon performance. As shown in the correlation heatmap (**Figure 3.1**), WBGT and dry bulb temperature were the most impactful environmental variables, but their correlations with marathon times were modest (around 0.10 for WBGT and DBT in some races). Given the relatively low correlations with other variables like solar radiation and dew point, WBGT emerged as the primary continuous measure of environmental stress, while flag conditions were used as a categorical measure for weather severity.

Figure 3.1: Correlation Between Weather Variables and Perforn



The bar plot showing correlations by race (**Figure 3.2**) highlighted the variation in these relationships across different marathons. For instance, Boston exhibited a 0.10 correlation between marathon times and DBT, while Duluth and Chicago had slightly higher correlations between WBGT and performance. These relatively small correlation values suggested that no single environmental variable had a dominant effect on marathon times, but WBGT consistently emerged as the most relevant continuous measure.



Moreover, the Tukey Post-Hoc Comparisons for Flag Conditions (**Table 3.1**) reinforced the importance of flag conditions in predicting marathon performance. Red flag conditions, representing the harshest weather, resulted in an average marathon time increase of 10.34 minutes compared to green flag conditions, while yellow flag conditions led to an average increase of 7.69 minutes. These differences were substantial across races, particularly in Duluth, where yellow flag conditions were most frequent. Interestly, the difference in performance times for runners in yellow flag vs red flag conditions were not significantly different (adjusted p = 0.32)

Tukey Post-Hoc Comparisons for Flag Conditions

Difference in Means	Lower Bound (95%)	Upper Bound (95%)	Adjusted P-value	Comparison
10.34	6.62	14.06	0.0000	Red-Green
−2.38	−4.24	−0.51	0.0060	White-Green
7.69	5.42	9.96	0.0000	Yellow-Green
−12.72	−16.49	−8.94	0.0000	White-Red
−2.65	−6.64	1.34	0.3200	Yellow-Red
10.07	7.71	12.42	0.0000	Yellow-White

When evaluating weather impact across age bins, older runners showed the greatest sensitivity to adverse weather. For instance, in the 65+ group, marathon times were significantly slower under red and yellow flag conditions, while younger age groups (such as 18-25) showed less variation in their times across flag categories (**Figure 2.1**). This difference highlights that flag conditions, particularly red and yellow flags, have a much more pronounced effect on older runners than on their younger counterparts.

In summary, Aim 3 demonstrated that WBGT and flag conditions had the most significant impacts on marathon performance. While the correlations were relatively modest, the substantial increases in marathon times under more severe flag conditions, particularly for older runners, confirmed the importance of WBGT as the key continuous variable and flag conditions as the primary categorical measure of weather impact.

Discussion and Conclusion

The exploratory analysis confirm well-established trends, such as the decline in performance with age, but also reveal nuances, particularly how men and women are differently affected by aging and environmental stressors. While both men and women show peak marathon performance in their late 20s to early 30s, the decline in performance with age is more pronounced in women, particularly in the 65+ age group. The growing gender gap in older runners suggests that physiological factors like muscle loss and thermoregulation may affect women more severely as they age. Notably, Duluth's outlier, where women seem to peak six years later than men, raises questions about whether this is linked to the race's summer heat, which might affect women's performance differently.

The environmental conditions, particularly WBGT and flag conditions, played a significant but relatively modest role in affecting performance. Duluth's high WBGT, for example, explains some of the slower marathon times compared to races held in cooler seasons, like Boston or New York. The results also showed that older runners were more sensitive to adverse weather.

Interestingly, our analysis showed no significant relationship between ozone levels and marathon performance. This could be due to generally lower ozone levels during the events or the brief exposure time for runners.

Future research could look into the physiological mechanisms driving performance declines, particularly the gendered responses to heat and age. As climate change intensifies, it will also be critical to examine how rising temperatures and increasingly erratic weather conditions may further impact marathon performance and the safety of runners.

References:

1. Ely, B. R., Cheuvront, S. N., Kenefick, R. W., & Sawka, M. N. (2010). Aerobic performance is degraded, despite modest hyperthermia, in hot environments. Med Sci Sports Exerc, 42(1), 135-41.
2. 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.
3. Kenney, W. L., & Munce, T. A. (2003). Invited review: aging and human temperature regulation. Journal of applied physiology, 95(6), 2598-2603.
4. 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.
5. Yanovich, R., Ketko, I., & Charkoudian, N. (2020). Sex differences in human thermoregulation: relevance for 2020 and beyond. Physiology, 35(3), 177-184.

Code appendix

```
knitr::opts_chunk$set(warning = FALSE,
                      message = FALSE,
                      echo = FALSE,
                      fig.align = "center")
# load libraries
library(tidyverse)
library(gt)
library(lubridate)
library(reshape2)
library(broom)


################## DATA IMPORT AND INITIAL CLEANING ##################

# read in datasets
course_record <- read.csv("~/Downloads/course_record-2.csv")
marathon_dates <- read.csv("~/Downloads/marathon_dates.csv")
aqi_vals <- read.csv("~/Downloads/aqi_values.csv")
project1 <- read.csv("~/Downloads/project1.csv")

# convert course record times to seconds
course_record <- course_record %>%
  mutate(CR_sec = as.numeric(hms(course_record$CR)))

# map race and sex
race_map <- c("0" = "B", "1" = "C", "2" = "NY", "3" = "TC", "4" = "D")
sex_map <- c("0" = "F", "1" = "M")

# adjust race mapping for duluth (grandma's marathon)
race_map.1 <- c("B" = "Boston", "C" = "Chicago", "NY" = "New York",
                "TC" = "Twin Cities", "D" = "Duluth (Grandma's)")


################## DATA MERGING ##################
# apply race and sex mappings to project1
project1 <- project1 %>%
  mutate(
    Race_standard = race_map[as.character(Race..0.Boston..1.Chicago..2.NYC..3.TC..4.D.)],
    Sex_standard = sex_map[as.character(Sex..0.F..1.M.)]
  )




# join project1 with course_record
merged_df <- project1 %>%
  left_join(course_record, by = c("Race_standard" = "Race", "Year",
                                  "Sex_standard" = "Gender"))

# remove unnecessary columns, calculate time in minutes
merged_df <- merged_df %>%
  select(-Race..0.Boston..1.Chicago..2.NYC..3.TC..4.D., -Sex..0.F..1.M.) %>%
  mutate(time_min = (CR_sec * (1 + (X.CR / 100))) / 60)

# rename columns for clarity
merged_df <- merged_df %>%
  rename(
    race = Race_standard,
```

```r
    year = Year,
    sex = Sex_standard,
    flag = Flag,
    age = Age..yr.,
    percent_cr = X.CR,
    dew_point = DP,
    rel_humid = X.rh,
    globe_temp = Tg..C,
    wb_temp = Tw..C,
    db_temp = Td..C,
    solar_rad = SR.W.m2,
    wind = Wind
  ) %>%
  mutate(race = race_map.1[race])

# create age bins
merged_df <- merged_df %>%
  mutate(age_bin = cut(age, breaks = c(0, 17, 25,35,45,55,65, Inf),
                       labels = c("< 18", "18-25", "26-35", "36-45",
                                  "46-55", "56-64", "65+")))

# ensure race names are consistent between marathon_dates and merged_df
marathon_dates <- marathon_dates %>%
  mutate(race = case_when(
    marathon == "NYC" ~ "New York",
    marathon == "Grandmas" ~ "Duluth (Grandma's)",
    T ~ marathon
  ))

# join with marathon_dates
marathon_dates <- marathon_dates %>%
  mutate(date = as.Date(date, format = "%Y-%m-%d"))

merged_df <- merged_df %>%
  left_join(marathon_dates, by = c("race", "year"))


######### MERGE AQI VAL INFO ##################

# standardize race names in aqi_vals
aqi_vals <- aqi_vals %>%
  rename(race = marathon) %>%
  mutate(
    race = case_when(
      race == "NYC" ~ "New York",
      race == "Grandmas" ~ "Duluth (Grandma's)",
      T ~ race
    ),
    date = as.Date(date_local, format = "%Y-%m-%d"),
    year = as.numeric(format(date, "%Y"))
  ) %>%
  select(-date_local)

# calculate average ozone ppm (8-hour avg)
avg_ppm <- aqi_vals %>%
  filter(units_of_measure == "Parts per million",
         sample_duration == "8-HR RUN AVG BEGIN HOUR") %>%
  group_by(race, year, date) %>%
  summarize(avg_ppm = mean(arithmetic_mean, na.rm = T)) %>%
```

```
  ungroup()

# join ppm data with merged_df, final join
merged_df <- merged_df %>%
  left_join(avg_ppm, by = c("race", "year", "date"))


########## EXPLORING MISSING DATA ##################

# weather related variables
weather_vars <- c("db_temp", "wb_temp", "rel_humid", "globe_temp", "solar_rad",
                  "dew_point", "wind", "WBGT")

# total observations in the data
total_observations <- nrow(merged_df)

# filter missing weather data
missing_weather <- merged_df %>%
  filter_at(vars(weather_vars), any_vars(is.na(.))) %>%
  group_by(year, race) %>%
  summarise(
    missing_count = n()  # number ofmissing rows for each year and race
  ) %>%
  mutate(missing_percentage = (missing_count / total_observations) * 100) %>%
  arrange(year, race)

# total missing data for all years and races combined
total_missing <- sum(missing_weather$missing_count)
total_missing_percentage <- (total_missing / total_observations) * 100

# convert year to character
missing_weather <- missing_weather %>%
  mutate(year = as.character(year))

# total summary row at the bottom of the table
missing_weather <- missing_weather %>%
  bind_rows(
    tibble(
      year = "Total",
      race = "",
      missing_count = total_missing,
      missing_percentage = total_missing_percentage
    )
  )

# table with percentage of missing data and bold years
missing_weather_table <- missing_weather %>%
  gt() %>%
  tab_header(
    title = "Table 0: Summary of Missing Weather Data by Year and Race",
    subtitle = "Missing Data Relative to Total Observations"
  ) %>%
  cols_label(
    year = "Year",
    race = "Race",
    missing_count = "Missing Data Count",
    missing_percentage = "Missing Data (%)"
  ) %>%
  fmt_number(
```

```
    columns = vars(missing_count),
    decimals = 0
  ) %>%
  fmt_number(
    columns = vars(missing_percentage),
    decimals = 2
  ) %>%
  tab_style(
    style = list(
      cell_text(weight = "bold")
    ),
    locations = cells_body(
      columns = vars(year)
    )
  ) %>%
  tab_style(
    style = list(
      cell_text(weight = "bold")
    ),
    locations = cells_body(
      rows = year == "Total"
    )
  )

missing_weather_table

###### DATA CLEANING; DROP NAs #################
cleaned_df <- merged_df %>%
  drop_na(db_temp, wb_temp, rel_humid, globe_temp,
          solar_rad, dew_point, wind, WBGT)


age_time_bp <- ggplot(cleaned_df, aes(x = age_bin, y = time_min, fill = sex)) +
  facet_wrap(~race) +
  geom_boxplot(outlier.size = 0.5, size = 0.35) +
  labs(x = "Age", y = "Marathon Time in Minutes",
       title = "Figure 1.1: Marathon Times by Age and Gender",
       fill = "sex") +
  scale_fill_manual(values = c("F" = "#BE398D", "M" = "#43A5BE")) +
  theme_classic(base_family = "serif") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "bottom")


age_time_bp

marathon_summary <- cleaned_df %>%
  group_by(age_bin) %>%
  summarize(
    # male count and percentage
    male_counts = paste0(n_male <- sum(sex == "M", na.rm = T),
                         " (", round(mean(sex == "M", na.rm = T) * 100, 1), "%)"),

    # female count and percentage
    female_counts = paste0(n_female <- sum(sex == "F", na.rm = T),
                           " (", round(mean(sex == "F", na.rm = T) * 100, 1), "%)"),

    # mean and sd for male times
    male_time = paste0(round(mean(time_min[sex == "M"], na.rm = T), 1),
                       " (", round(sd(time_min[sex == "M"], na.rm = T), 1), ")"),
```

```r
    # mean and sd for female times
    female_time = paste0(round(mean(time_min[sex == "F"], na.rm = T), 1),
                         " (", round(sd(time_min[sex == "F"], na.rm = T), 1), ")"),

    # total number of participants in each age group
    n_total = n()
  ) %>%
  ungroup()


marathon_summ_table <- marathon_summary %>%
  gt() %>%
  tab_header(
    title = "Table 1.1: Marathon Performance Summary by Age Group and Sex"
  ) %>%
  cols_label(
    age_bin = "Age Group",
    male_counts = "Male (n, %)",
    female_counts = "Female (n, %)",
    male_time = "Male Marathon Time",
    female_time = "Female Marathon Time",
    n_total = "n"
  ) %>%
  cols_align(
    align = "center",
    columns = everything()
  ) %>%
  tab_style(
    style = cell_text(weight = "bold"),
    locations = cells_column_labels(everything())
  ) %>%
  fmt_number(
    columns = c("n_total"),
    decimals = 0
  ) %>%
  tab_footnote(
    footnote = "Marathon time is presented as mean (SD) in minutes.",
    locations = cells_column_labels(columns = c("male_time", "female_time"))
  )

marathon_summ_table


# anova model with interaction between sex, age bin, and flag
anova_model_interaction <- aov(time_min ~ sex * age_bin * flag, data = cleaned_df)

# anova summary
anova_summary_interaction <- summary(anova_model_interaction)

# tukey's post-hoc test on sex, age_bin, and flag
tukey_interaction <- TukeyHSD(anova_model_interaction)

# extract anova results
anova_results_interaction <- tidy(anova_model_interaction)

# anova gt table
anova_table_interaction <- anova_results_interaction %>%
  gt() %>%
  tab_header(
```

```r
    title = "Table 1.2: ANOVA Results for Marathon Times by Flag Condition, Gender, and Age Bin"
  ) %>%
  cols_label(
    term = "Source",
    df = "Df",
    sumsq = "Sum Sq",
    meansq = "Mean Sq",
    statistic = "F-value",
    p.value = "Pr(>F)"
  ) %>%
  fmt_number(
    columns = c(sumsq, meansq),
    decimals = 0
  ) %>%
  fmt_number(
    columns = c(statistic, p.value),
    decimals = 3
  ) %>%
  tab_options(
    table.font.size = "medium",
    heading.title.font.size = 16
  ) %>%
  cols_align(
    align = "center",
    columns = everything()
  )


anova_table_interaction

# tukey post-hoc test
tukey_flag <- TukeyHSD(anova_model_interaction, "flag")
tukey_sex <- TukeyHSD(anova_model_interaction, "sex")
tukey_age <- TukeyHSD(anova_model_interaction, "age_bin")

# convert tukey results to data frames
tukey_results_flag <- as.data.frame(tukey_flag$flag)
tukey_results_sex <- as.data.frame(tukey_sex$sex)
tukey_results_age <- as.data.frame(tukey_age$age_bin)

# add comparison names to Tukey results
tukey_results_flag$Comparison <- rownames(tukey_results_flag)
tukey_results_sex$Comparison <- rownames(tukey_results_sex)
tukey_results_age$Comparison <- rownames(tukey_results_age)

tukey_sex_table <- tukey_results_sex %>%
  gt() %>%
  tab_header(
    title = "Table 1.3: Tukey Post-Hoc Test Results for Sex",
    subtitle = "95% Family-Wise Confidence Level"
  ) %>%
  cols_label(
    Comparison = "Sex Comparison",
    diff = "Difference",
    lwr = "Lower Bound",
    upr = "Upper Bound",
    `p adj` = "Adjusted p-value"
  ) %>%
  fmt_number(
```

```r
    columns = c("diff", "lwr", "upr", "p adj"),
    decimals = 3
  )

tukey_sex_table


# average percent_cr for each age and sex
avg_percent_age <- cleaned_df %>%
  group_by(age, sex) %>%
  summarise(avg_percent_cr = mean(percent_cr, na.rm = TRUE), .groups = 'drop')

# calculate the average time each age, sex, and race
avg_time_age <- cleaned_df %>%
  group_by(age, sex, race) %>%
  summarise(avg_time = mean(time_min, na.rm = TRUE), .groups = 'drop')

# min percent for each age and sex
min_percent_by_sex <- avg_percent_age %>%
  group_by(sex) %>%
  summarise(min_percent_cr = min(avg_percent_cr),
            min_age = age[which.min(avg_percent_cr)],
            .groups = 'drop')


# minimum average time and corresponding age for each sex and race
min_time_by_sex_race <- avg_time_age %>%
  group_by(sex, race) %>%
  summarise(min_time = min(avg_time),
            min_age = age[which.min(avg_time)],
            .groups = 'drop')


# plot showing average time, showing minimum age by sex
age_time_avg <- ggplot(avg_time_age, aes(x = age, y = avg_time, color = sex)) +
  geom_point(shape = "square", size = .5) +
  geom_line(size = .25) +
  geom_hline(data = min_time_by_sex_race, aes(yintercept = min_time, color = sex), linetype = "dashed") +  # H
  geom_text(data = min_time_by_sex_race,
            aes(x = min_age, y = min_time, label = paste0(sex, " Min Age: ", min_age)),
            hjust = -1.6, vjust = 0, color = "black", size = 3.25, family = "serif") +
  facet_wrap(~race) +  # Facet by race
  labs(x = "Age", y = "Average Marathon Time (minutes)",
       title = "Figure 1.2: Average Marathon Performance by Age and Race",
       color = "Sex") +
  scale_color_manual(values = c("F" = "#BE398D", "M" = "#43A5BE")) +
  theme_classic(base_family = "serif") +
  theme(legend.position = "bottom")


age_time_avg


tukey_age_table <- tukey_results_age %>%
  gt() %>%
  tab_header(
    title = "Table 1.3: Tukey Post-Hoc Test Results for Age Bins",
    subtitle = "95% Family-Wise Confidence Level"
  ) %>%
```

```r
  cols_label(
    Comparison = "Age Bin Comparison",
    diff = "Difference",
    lwr = "Lower Bound",
    upr = "Upper Bound",
    `p adj` = "Adjusted p-value"
  ) %>%
  fmt_number(
    columns = c("diff", "lwr", "upr", "p adj"),
    decimals = 3
  )

tukey_age_table

# summarize the weather data by race
weather_summary_race <- cleaned_df %>%
  group_by(race) %>%
  summarize(
    wb_temp = paste0(round(median(wb_temp, na.rm = TRUE), 1), " (",
                     round(quantile(wb_temp, 0.25, na.rm = TRUE), 1), ",",
                     round(quantile(wb_temp, 0.75, na.rm = TRUE), 1), ")"),
    db_temp = paste0(round(median(db_temp, na.rm = TRUE), 1), " (",
                     round(quantile(db_temp, 0.25, na.rm = TRUE), 1), ",",
                     round(quantile(db_temp, 0.75, na.rm = TRUE), 1), ")"),
    rel_humid = paste0(round(median(rel_humid, na.rm = TRUE), 1), " (",
                       round(quantile(rel_humid, 0.25, na.rm = TRUE), 1), ",",
                       round(quantile(rel_humid, 0.75, na.rm = TRUE), 1), ")"),
    solar_rad = paste0(round(median(solar_rad, na.rm = TRUE), 1), " (",
                       round(quantile(solar_rad, 0.25, na.rm = TRUE), 1), ",",
                       round(quantile(solar_rad, 0.75, na.rm = TRUE), 1), ")"),
    wind = paste0(round(median(wind, na.rm = TRUE), 1), " (",
                  round(quantile(wind, 0.25, na.rm = TRUE), 1), ",",
                  round(quantile(wind, 0.75, na.rm = TRUE), 1), ")"),
    WBGT = paste0(round(median(WBGT, na.rm = TRUE), 1), " (",
                  round(quantile(WBGT, 0.25, na.rm = TRUE), 1), ",",
                  round(quantile(WBGT, 0.75, na.rm = TRUE), 1), ")")
  )

# create the gt table for weather summary by race
weather_gt_table <- weather_summary_race %>%
  gt() %>%
  tab_header(
    title = "Table 2.1: Weather Summary by Race",
    subtitle = "Median (IQR) of Weather Variables"
  ) %>%
  cols_label(
    race = "Race",
    wb_temp = "WBT (°C)",
    db_temp = "DBT (°C)",
    rel_humid = "RH (%)",
    solar_rad = "SR (W/m²)",
    wind = "Wind (m/s)",
    WBGT = "WBGT (°C)"
  ) %>%
  tab_style(
    style = cell_text(weight = "bold", align = "center"),
    locations = cells_column_labels(everything())
  ) %>%
  tab_options(
```

```
    table.font.size = "small",
    heading.title.font.size = 16,
    heading.subtitle.font.size = 12
  ) %>%
  tab_footnote(
    footnote = "WBT: Wet Bulb Temp, DBT: Dry Bulb Temp, RH: Relative Humidity,
    SR: Solar Radiation, WBGT: Wet Bulb Globe Temp",
    locations = cells_column_labels(columns = c("wb_temp", "db_temp",
                                               "rel_humid", "solar_rad",
                                               "wind", "WBGT"))
  ) %>%
  tab_style(
    style = cell_text(size = px(10)),
    locations = cells_footnotes()
  )


weather_gt_table

# revise flag colors with more distinct and visible names
flag_colors <- c("White" = "#C1B8FF",
                 "Green" = "#005A36",
                 "Yellow" = "#FFEA65",
                 "Red" = "#BE398D")

# boxplot with age bins and flag color
boxplot_age_bin <- ggplot(cleaned_df, aes(x = age_bin, y = percent_cr, fill = flag)) +
  geom_boxplot(outlier.size = 0.5, size = 0.25)+
  facet_wrap(~race, ncol = 2) +  # Facet by race
  scale_fill_manual(values = flag_colors) +
  labs(title = "Figure 2.1: Marathon Performance by Age Bin and Flag Condition",
       x = "Age Bin",
       y = "Marathon Time (% Course Record)",
       fill = "Flag Condition") +
  theme_classic(base_family = "serif") +
  theme(
    panel.grid.major = element_line(size = 0.25),
    legend.position = "bottom"
  )

boxplot_age_bin

# summarize flag
flag_summary <- cleaned_df %>%
  group_by(race, flag) %>%
  summarize(flag_count = n()) %>%
  ungroup() %>%
  group_by(race) %>%
  mutate(flag_percent = flag_count / sum(flag_count))


# proportion of flags for each race bar plot
prop_flag <- ggplot(flag_summary, aes(x = flag, y = flag_percent, fill = flag)) +
  geom_bar(stat = "identity", width = 0.6, color = "black") +
  geom_text(aes(label = scales::percent(flag_percent, accuracy = 1)),
            position = position_stack(vjust = 0.5), size = 3, family = "serif") +
  labs(x = "Flag Condition",
       y = "Proportion of Flag Conditions",
       title = "Flag Conditions by Race",
```

```
       family = "serif") +
  scale_fill_manual(values = flag_colors) +
  facet_wrap(~race) +
  coord_flip() +
  theme_classic(base_family = "serif") +
  theme(legend.position = "none")

prop_flag


ppm_percent_cr <- ggplot(cleaned_df, aes(x = avg_ppm, y = percent_cr, color = sex)) +
  geom_point(alpha = 0.1) +
  geom_smooth(se = FALSE, size = 1) +
  facet_wrap(~age_bin, scales = "free_y", nrow = 4) +
  labs(title = "Figure 2.3: Air Quality (Ozone PPM) vs Marathon Time by Sex & Age Group",
       x = "Ozone PPM",
       y = "Marathon Time (% Percent CR)",
       color = "Sex") +
  scale_color_manual(values = c("F" = "#BE398D", "M" = "#43A5BE")) +
  theme_classic(base_family = "serif") +
  theme(legend.position = "bottom")

ppm_percent_cr

cor_vars <- cleaned_df %>%
  select(time_min, dew_point, rel_humid, globe_temp, wb_temp, db_temp,
         solar_rad, wind, WBGT)

cor_matrix <- cor(cor_vars, use = "complete.obs")
melted_cor_matrix <- melt(cor_matrix)

cor_plot <- ggplot(data = melted_cor_matrix, aes(x = Var1, y = Var2, fill = value)) +
  geom_tile(color = "white") +
  scale_fill_gradient2(low = "#090C9B", high = "#BE398D", mid = "white",
                       midpoint = 0, limit = c(-1, 1), space = "Lab",
                       name="Correlation") +
  labs(title = "Figure 3.1: Correlation Between Weather Variables and Performance") +
  theme_minimal(base_family = "serif") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1), legend.position = "bottom")

cor_plot

race_colors = c("Boston" = "#005A36", "New York" = "#090C9B", "Chicago" = "#BE398D",
                "Duluth (Grandma's)" = "#43A5BE", "Twin Cities" = "#C1B8FF")

# group by race, calculate correlations between weather and marathon time
correlations_by_race <- cleaned_df %>%
  group_by(race) %>%
  summarize(
    wb_temp_corr = cor(time_min, wb_temp, use = "complete.obs"),
    db_temp_corr = cor(time_min, db_temp, use = "complete.obs"),
    rel_humid_corr = cor(time_min, rel_humid, use = "complete.obs"),
    solar_rad_corr = cor(time_min, solar_rad, use = "complete.obs"),
    wind_corr = cor(time_min, wind, use = "complete.obs"),
    WBGT_corr = cor(time_min, WBGT, use = "complete.obs")

  ) %>%
  gather(key = "weather_var", value = "correlation", -race)
```

```r
# plot the correlations by race
corr_race_plot <- ggplot(correlations_by_race, aes(x = weather_var, y = correlation, fill = race)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_hline(yintercept = 0, linetype = "dashed", color = "black", size = 0.5) +
  scale_fill_manual(values = race_colors) +
  labs(
    title = "Figure 3.2: Correlations Between Weather Variables and Marathon Times",
    x = "Weather Variable",
    y = "Correlation with Marathon Time"
  ) +
  theme_classic(base_family = "serif") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "bottom")


corr_race_plot


# tukey table for flag condition
tukey_table_flag <- tukey_results_flag %>%
  gt() %>%
  tab_header(
    title = "Tukey Post-Hoc Comparisons for Flag Conditions"
  ) %>%
  cols_label(
    Comparison = "Comparison",
    diff = "Difference in Means",
    lwr = "Lower Bound (95%)",
    upr = "Upper Bound (95%)",
    `p adj` = "Adjusted P-value"
  ) %>%
  fmt_number(
    columns = c(diff, lwr, upr),
    decimals = 2
  ) %>%
  fmt_number(
    columns = `p adj`,
    decimals = 4
  ) %>%
  tab_options(
    table.font.size = "medium",
    heading.title.font.size = 16
  ) %>%
  cols_align(
    align = "center",
    columns = everything()
  )

tukey_table_flag
```