

Boston Marathon Run Times

Ashwath Mohan

Nathaniel Medina

Nicolas Pappas

Rick Yang

University of California Berkeley

Abstract

With increased worldwide public health awareness and emphasis on healthy lifestyles, marathons have become a very common activity in recent years. The Boston Marathon is notorious for having one of the toughest qualifying standards for a marathon and, as a result, attracts competitors from all over the world. Recent studies have claimed that the average run time among participants is getting slower year by year. Is this because humans are getting slower? Or are there other factors at play? By analyzing 14 consecutive years of Boston Marathon data, we aim to answer these questions.

1 Introduction

The Boston Marathon is an iconic annual event in the United States. We thought it would be both meaningful and fun to investigate publicly available Boston Marathon data, with the intention of finding deeper insights and trends about this annual event. In addition, our group members enjoy running and have goals to complete marathons in the future. One topic we aim to focus on is the number of participants. We will split this into smaller subcategories such as age, gender, and country of origin. Some questions we hope to answer are: Which countries have the most participants? What

is the most common age group? Which gender tends to participate more? How many returning runners are there every year?

Another topic we would like to analyze is the average run time. We will observe the run times of every participant to get a baseline average for each year. After that, we will look at the average run time based on age, gender, and country in each year.

Ultimately, our goal is to see whether or not age, gender, and country have an effect on performance. By analyzing these subcategories, we will be able to address whether or not times have been getting slower over the years and determine if there are any underlying factors that could have caused this trend.

2 Hypothesis

An increase in public health awareness and the importance of living a healthy lifestyle has caused the focus of a marathon to lean more towards the social aspect. More participants are running in the Boston Marathon as the race increases in popularity and competitiveness around the world. The global prestige from competing in this race is encouraging participants from various countries to compete in the Boston Marathon. Participants run the race for the

experience of participating in a world-known famous event without trying to have the best run time. Because of these factors, our hypothesis is that the average run time among all participants has increased over the years.

3 Related Works

Previously, there have been many works related to the Boston Marathon data. One project examined the people, performance, and other topics arising out of the Boston Marathon in 2017. A similar analysis was done on the finishers of the Boston Marathon from 2016. Another project analyzed Boston Marathon data across 2015, 2016, 2017. Many of these projects investigate data from recent years, but they neglect to consider data from over a decade ago. This is where our analysis differs from the rest, since we strive to analyze Boston Marathon data across 14 years in order to recognize deeper insights and trends. By doing this, we will be able to see how the people and performance relating to the race has changed over the years.

4 Data

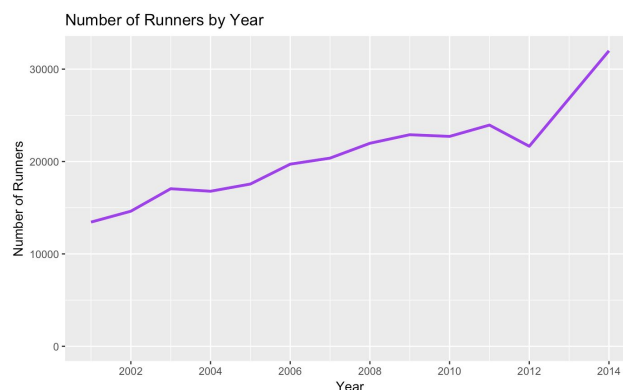
The data set we used was retrieved from Github. It contains detailed information about all the runners that participated in the Boston Marathon across 14 years. Within each year, multiple columns of data are presented for each runner. This includes information about his or her name, city of origin, country of origin, gender, age, race time, race ranking, and more. With such a

wide range of data, we were able to make insightful discoveries about the runners and the Boston Marathon itself.

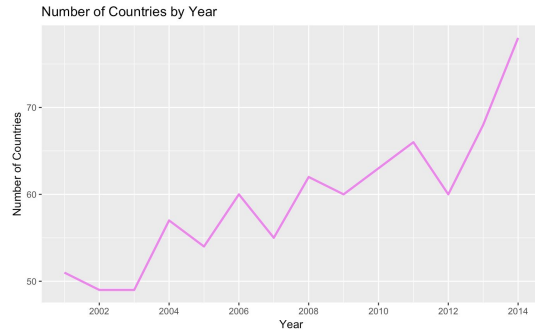
To start our data analysis, we began by aggregating 14 years of data from separate csv files. This presented a challenge because not all of the datasets had the same format or columns. After combining the data, we proceeded by selecting relevant columns of data, creating additional columns (such as the Years column), changing the data type for certain columns, and removing N/A values. Then, our data was in a cleaned format that would allow us to conduct our analysis.

5 Visualizations

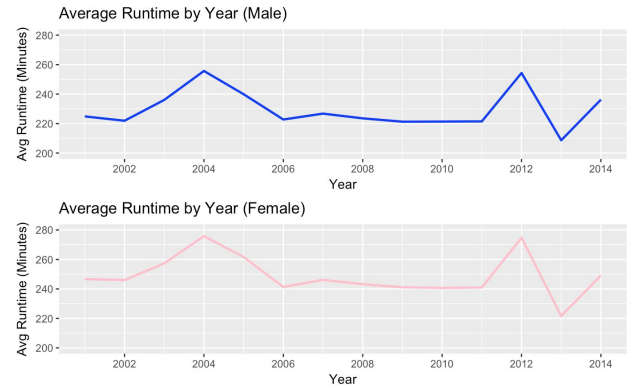
5.1 Analysis of Participants



5.1.1 The number of runners is consistently increasing.

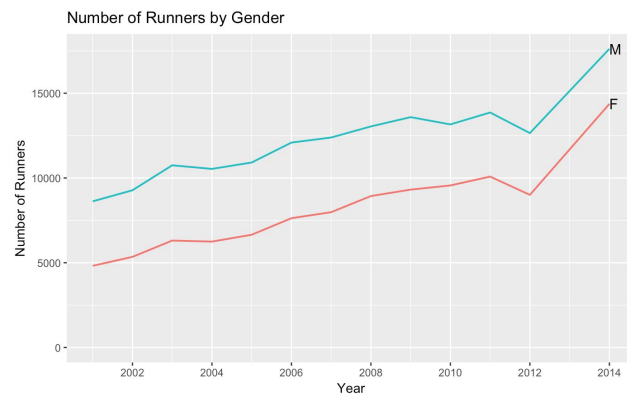


5.1.2 The number of countries consistently increases from 2001 to 2014, showing that the race gained popularity from international runners.

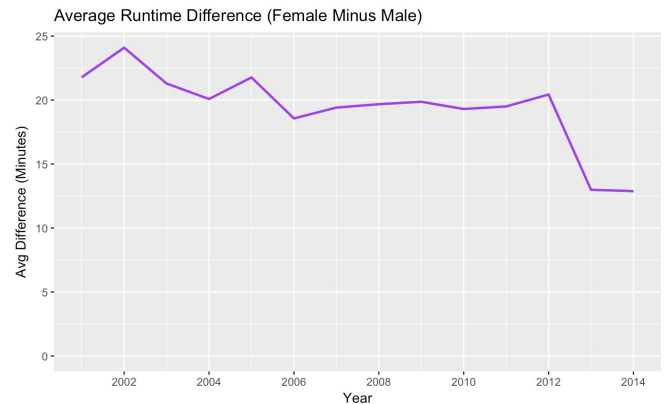


5.2.2 On average, men run faster than women.

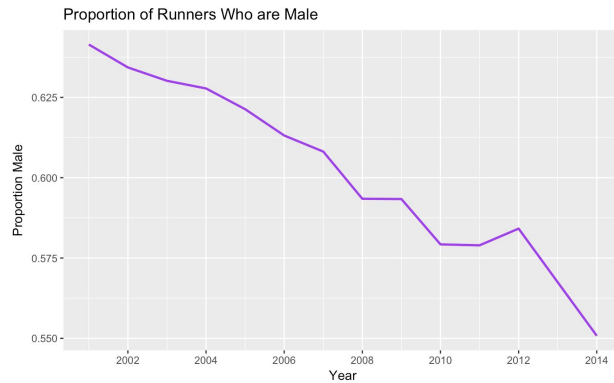
5.2 Gender Analysis



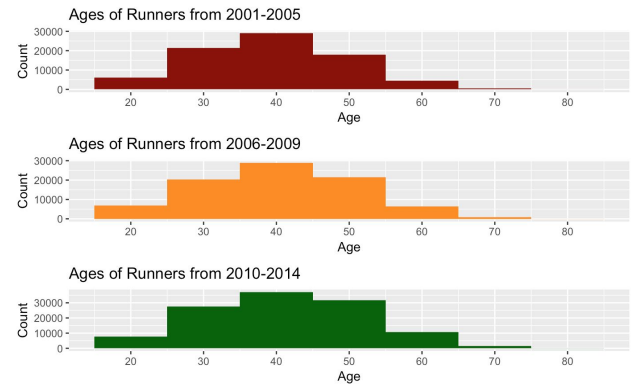
5.2.1 Every year has more male runners than female runners. The gap between male and female runners is closing.



5.2.3 Although men run faster on average, the runtime difference between genders is decreasing. Possibly due to the decrease in male runners participating.

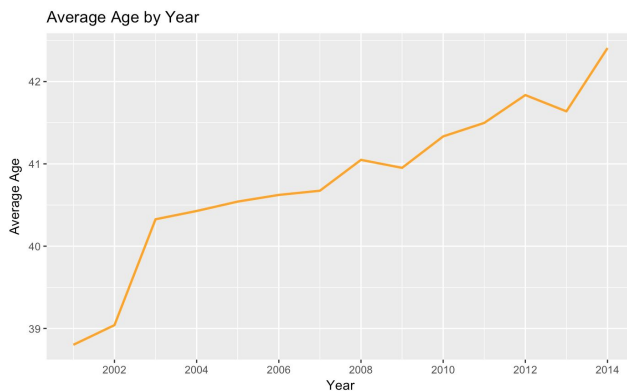


5.2.4 Fewer male runners are participating in the Boston Marathon in proportion to female runners.

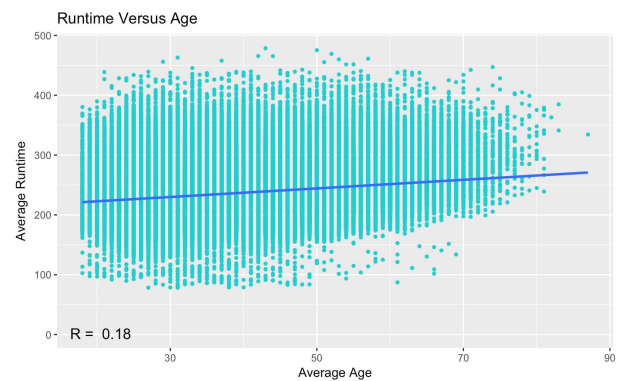


5.3.2 Further supports the idea that runners are getting older. This change is most easily seen in the Age bucket labeled “50” years across all three histograms.

5.3 Age Analysis

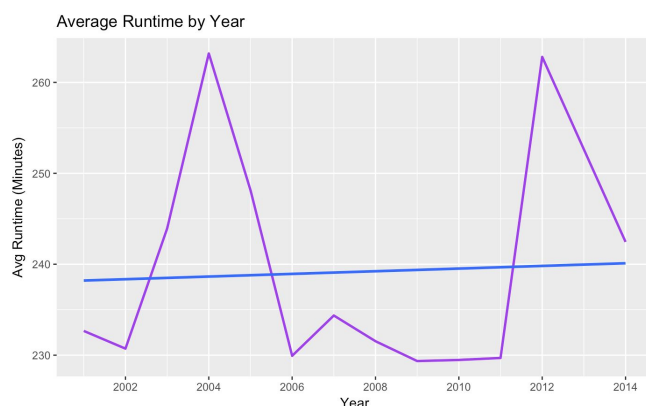


5.3.1 Average age of runners is consistently getting older.



5.3.3 There is a direct relationship between age and run time. Older runners have greater run times. Note: the blue line represents the regression line, which also shows a long-run increase in average age.

5.4 Average Run Time Analysis



5.4.1 The purple line represents the average runtime per year. There are spikes in average runtimes in the years 2004 and 2012 due to abnormally high weather conditions, causing many to suffer from heat exhaustion and dehydration (see references). To get a better understanding of the overall trend, we used a regression line (in blue) which shows that the average run times are slowly increasing from 2002 to 2014.

6 Conclusions

Ultimately, our hypothesis that average runtime increases over time was proven correct, but not as significantly as we thought it would. The increased average age of runners and the closing gender gap disparities provide support for this. However, it is important to be aware that there could be confounding factors that influence the average runtime, such as increasing severity of weather conditions. Furthermore, the long term average runtime increase was rather small, and not significant enough to provide insights into more detailed conclusions. If we had

runtime data on 100 years, for example, we could obtain a more accurate trend line. But despite the small increase in runtime we observed, our hypothesis could be supported by the fact that average runner ages are increasing and the gender gap differences are narrowing. This reveals more about societal changes in that people are willing to run even though they are getting older and there is a continuing trend of gender equality in the races.

7 Final Thoughts

Throughout our analysis, we were convinced that average runtimes would see a significant increase throughout the 12 year analysis. However, since we did not see a significant increase, we were left wondering how we could make our analysis better. We came to the conclusion that more years would be needed to achieve a better trendline to rule out confounding factors. Moving forward, we would like to explore other datasets that include information from longer date ranges, such as 1950 to present time. This would require more data cleaning processes, but could illustrate more long term changes in demography and physiography. Additionally, the spikes in the data were due to heatwaves that caused runners to suffer from dehydration. It would be interesting to overlay weather data compared to runtimes to see if weather has an impact on each year's average runtime.

8 References

Github:

<https://github.com/llimllib/bostonmarathon>

2004 marathon: 1,100 runners suffered dehydration

http://archive.boston.com/sports/specials/marathon/articles/2004/04/20/heat_was_too_much_to_bear/

2012 marathon: 2,100 runners suffered from dehydration

http://archive.boston.com/sports/marathon/articles/2012/04/17/boston_marathon_runners_suffer_heat_related_ailments_in_record_breaking_temperatures/

Related Works:

The Runners of the Boston Marathon 2017:
A Data Analysis of the 2017 Boston Marathon

<https://medium.com/running-with-data/the-runners-of-the-boston-marathon-2017-56d2db7326c1>

Finishers Boston Marathon 2015, 2016, 2017

<https://www.kaggle.com/rojour/boston-results>

A Data Analysis of the 2016 Boston Marathon Finishers

<https://susanli2016.github.io/Boston-Marathon/>

