## Abstract

Participation in competitive running has increased dramatically in the past two decades. With an influx of new runners, new performance comparison metrics have been developed. One such metric is the age graded run time. Simultaneously, the rapid expansion of the Internet has resulted in massive quantities of data now freely available online. Given the appropriate tools, acquiring running data stored on a website is a trivial task. In this paper, we analyze the performance of runners in the Cherry Blossom Ten Mile Run held annually in Washington, D.C. Performance data is acquired by scraping the Cherryblossom.org race result archives for 1999 to 2012. We compare age-normalized race results for 1999 and 2012 men and women. Results show that 1999 runners performed better when using age grading. Furthermore, men typically outperformed women in both years considered.

## Introduction

The sport of running continues to grow rapidly. While major events like the Boston Marathon have held races for more than 100 years, the modern-day sport of running has seen a significant expansion of runner participation since 1995. Overall participation in races around the United States expanded from seven million in 1995 to almost 16 million in 2012 (Running USA 1).

The Cherry Blossom Ten Mile Run in Washington, D.C. represents a microcosm of the upward participation trend in running. Between the period of 1999 and 2012, participation has tripled. Runners of all ages and skill levels participate in the race each year.

The objective of any runner is to minimize his or her overall run time. However, age and run times have typically displayed a positive relationship. In order to account for this degradation in performance as participants age, run times are often normalized by the world record time for a given age or age group. This process, known as age grading, is used as a performance standard in competitive running (Nolan, et. al 2).

In this paper, we analyze race results from the Cherry Blossom Ten Mile Run. Results for the 1999 to 2012 period are accessible in HTML tables at Cherryblossom.org. While we explore overall trends and runner characteristics, our objective is to compare the distribution of age-normalized run times for 1999 and 2012. Additionally, we compare gender performance using the same age-normalized run times for both years.

We will review available literature on the topic of running performance and explore overall participation and performance trends from 1999 to 2012. We will discuss the methods used to compare 1999 to 2012 results for men and women and compare results. Our paper will conclude with a discussion of the applications of performance modeling in the increasingly data-driven world of personal fitness.

## Literature Review

Analyzing running performance is a common topic in medical, sports, and evolutionary journals. These studies have been conducted in order to more accurately determine where the human ability for long-distance running originated, how bodies are predisposed to running, and the impact of diet and training on performance.

One of the leading researchers on the subject of runner performance, Dr. Owen Anderson, showed both hereditary and environmental effects contribute to overall runner performance. However, no definitive proof could be given as to the level of impact each had on a runner’s performance (Anderson 3).

Further studies have shown that physical predisposition and genetics are major factors affecting running ability. Despite training and environmental factors, it is never enough to fully overcome physiology and the detrimental effects aging has on running ability (Tucker 4).

Performance in road races and competitive running is bounded by physiological factors. No matter how much diet or training is conducted, harm will come to the athlete who tries to go beyond these limits (Tucker 5,6). This phenomenon is seen in competitive running as a runner ages. Genetic predisposition for efficient oxygen recycling and higher anaerobic thresholds have also been found to enhance runner performance in other studies (Lorenz, et. al 7).

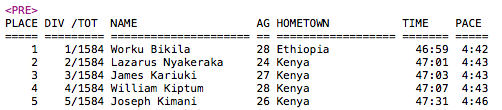
Ultimately, these studies show the human body has limitations. While there is a net performance gain over time, physiological factors limit human performance. We explore a way to control for age in our comparisons by implementing an age graded run time.

## Background and Methods

### Data Acquisition Summary

While data acquisition commonly presents itself in the form of reading a flat file, additional measures have to be taken to acquire data from embedded tables on a website. Formatting is often a concern, especially because website data are often surrounded by JavaScript and HTML markers. Luckily, the majority of data required for our analysis came in the form of readily accessible HTML tables stored at different end points on Cherryblossom.org. An example of one such table is seen in Table 1.

Table 1: Raw HTML Table for 1999 Runners



Data acquisition from websites, often called “web scraping,” is an iterative process. To prepare for data analysis, 28 tables of race results were acquired from Cherryblossom.org. Each table of data represents a year and gender combination. Each table is accessed via a unique web address. The formatting of HTML tables is inconsistent from year to year. In 1999, finish times were represented by a TIME variable. However, in other years, three time variables were presented. Additionally, messy HTML provided for minor difficulties in acquiring 2009 men’s race results. We also account for one unrealistic run time and 51 unrealistic or unreported ages.

Table 2: Formatted Dataset for Analysis

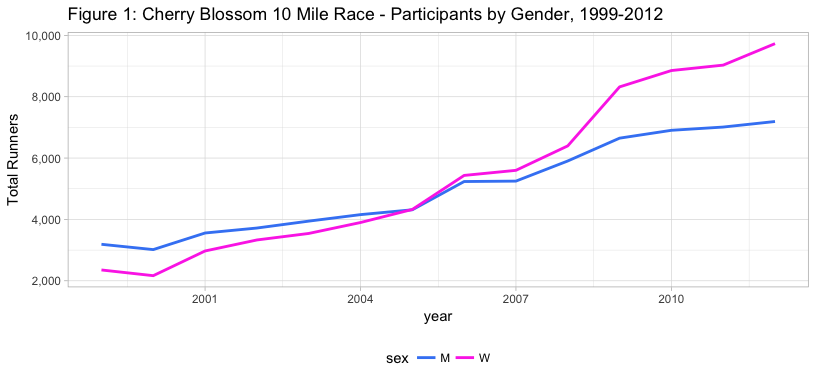
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| year | sex | name | home | age | runTime |
| 2012 | W | Neeti Doshi | Chapel Hill NC | 26 | 134.68333 |
| 2010 | M | Cory Plunkett | Clinton MD | 26 | 93.45000 |
| 2011 | W | Bernadette Domingo | Alexandria VA | 38 | 93.56667 |
| 2000 | W | Jill Rafano | Alexandria VA | 35 | 96.60000 |
| 2012 | M | Charles Dunbar | West Grove PA | 60 | 99.91667 |

Table 2 presents a snapshot of the dataset used for analysis. Sex, name, and home are all strings that provide the sex, first and last name and origin of a given runner. The three variables we are most interested in are year, age and runTime. These are all numeric variables. Year represents the calendar year the Cherry Blossom race was held. Age and runTime represent the age in years of the runner and their associated finish time in minutes.

Ultimately, we acquired 28 tables of race results for the 1999 to 2012 time period. While our main objective is to explore differences in only the 1999 and 2012 races, critical contextual information will be presented based on trends over the 14-year period from 1999 to 2012 prior to analysis results.

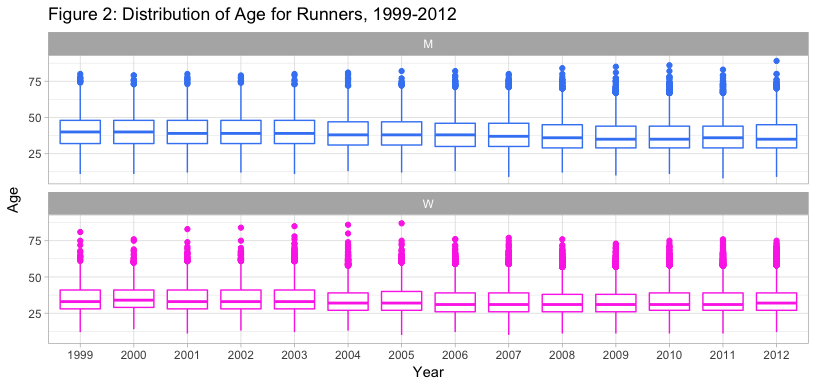
### Exploration of Cherry Blossom Ten Mile Race Data, 1999-2012

As stated previously, participation in the Cherry Blossom Ten Mile Run has grown rapidly. While participation by men and women both grew substantially, participation by women played the greatest role in the expansion of the field. Women’s participation comprised a minority of 42% in 1999. However, by 2012, women had exchanged positions with men at 57.5% of all participants. The two genders were at parity levels of participation in 2005, and women have since added more than 5,000 participants.

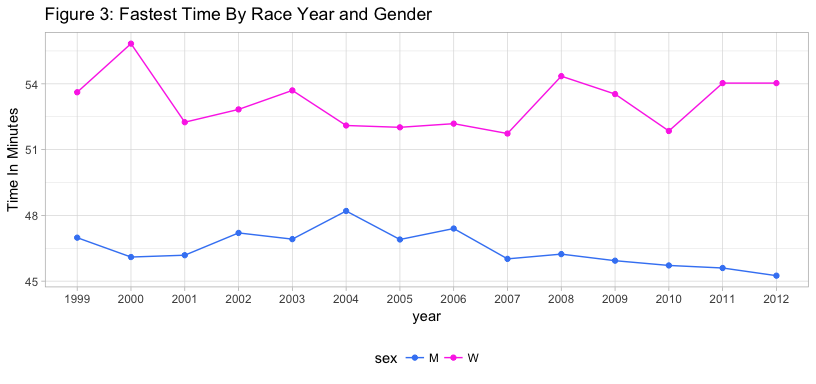


The age of male runners has declined since 1999, with an average age of 40.3 in 1999, declining to a minimum of 37.8 in 2008. The standard deviation of male ages has consistently hovered between 10 and 12 years.

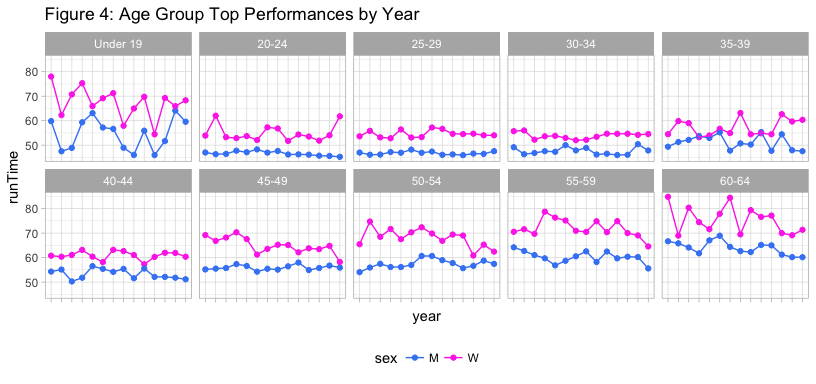
Women participating in the race, however, have consistently averaged between 33 and 36 years old, without any consistent trend year over year. Standard deviation of age for women is about nine years.



The finish time for the best male participant for this ten-mile event has improved consistently from 2008 to 2012, with times approaching 45 minutes. Based on the correlation between age and performance, the improvement in men’s times could be a result of the decrease in the average age for men. Women’s times have been more static, with the exception of a four-year run of times from 2004 to 2007.

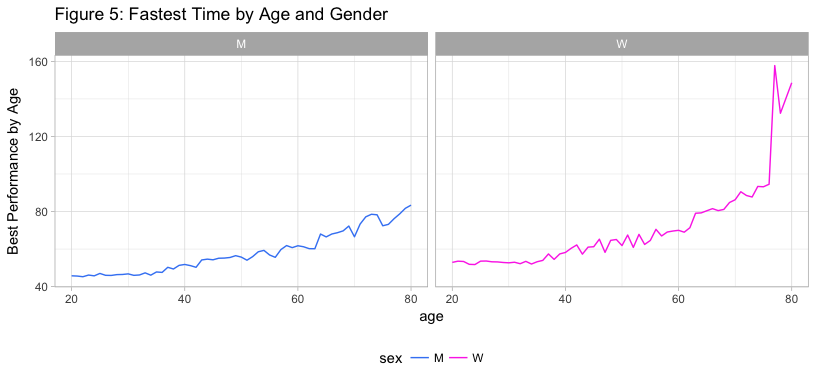


Examination of race finish times by age groups and gender show several general consistencies. The first is the consistency in most age groups from year to year of the race. Most age groups can be described as a “random walk” around a median race time, with the exception of the 20-24 year segment for men showing a consistent reduction of race finishing times year over year. The Under 19 segment shows a high degree of variation in the winning runner’s time. Some degree of this variability might be explained by the number of participants in the segment- consistently less than 100 in each gender- as well as progression of these runners into the more elite 20-24 age group. The two age divisions for ages 65 and over are omitted from the visualization.



There may also be some inference made about sudden changes in the winning times for age groups. The increase of the winning time in the 30-34 age group in the 2011 and 2012 races could signal a key competitor has aged out of the group. Indeed, we see the 35-39 age group for men taking a sudden drop in winning time in 2011 and 2012. While the race size overall is large, there could be a few elite competitors that set the pace for most of the age groups.

Examination of Figure 5 allows for another inference regarding the relationship between age and run time. The visualizations show the fastest running time for every age between 20 and 80 by gender. Excluding the youngest and oldest participants from our research, running times for men peak in the early twenties, with steadily increasing times for the rest of life. Women peak closer to thirty, with times slowing through their thirties and forties. Both genders’ run times display non-linear relationships when considering age, especially as age increases.

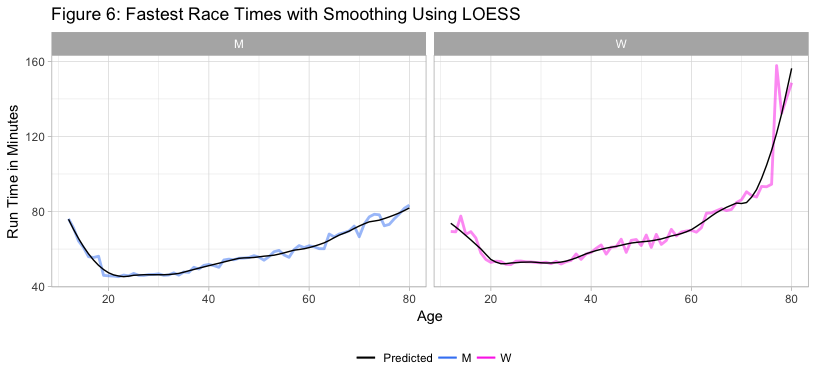


### Controlling for Age in Competitive Race Results

The relationship between age and performance is a natural phenomenon known well by competitive runners. To account for this phenomenon, age grading is used for ranking runners’ performance based on their age. In fact, age graded performance is more frequently used for runner comparisons. Runnersworld.com describes age grading as “a way of placing all race participants on a level playing field, regardless of age or gender. (Runnersworld 3)” In age grading, a runner’s time is normalized based on the world record time for his or her age. To normalize run times, analysis typically involves dividing raw run times by the world record for a given age. As an example, a runner who is 33 with a finish time of 75 minutes would have a standardized time of (75/57) 1.32 given a record time of 57 minutes for his age category. We use the terms age grading and normalization interchangeably for the remainder of this paper.

Given overwhelming visual evidence of the positive relationship between age and running performance, we implement an age grading method to compare the 1999 results to the 2012 race results for the Cherry Blossom Ten Mile Run. Top runners in the Cherry Blossom races have run very close to world records. Therefore, we use the fastest times for each age in the dataset containing race results from 1999 to 2012. Typically, we would use the actual fastest run times for each age to normalize raw run times. However, year-to-year fluctuations of fastest times given age call for a different approach.

To account for run time fluctuations, we gather the fastest times for each age predicted by a LOESS model. LOESS is a non-parametric local regression method that uses subsets of data to fit individual regression models across an explanatory variable range. This method allows for complex curve fitting. In our case, it takes some of the noise out of the run times we use for age grading. We fit the LOESS model to the dataset containing the fastest run times and age, using run time as the response. We subsequently predict fastest run times for each age given our fitted model. The resulting predictions then represent smoothed fastest times to be used for normalization. The predictions for fastest times given age are represented by the black line for both genders in Figure 6. LOESS has effectively smoothed the fastest run times used for age grading.



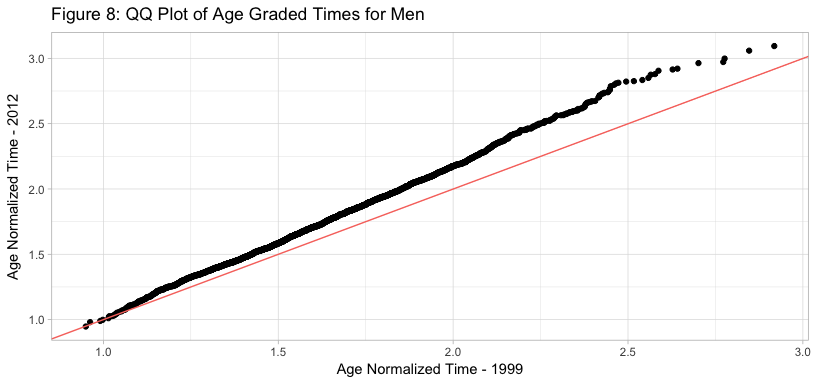
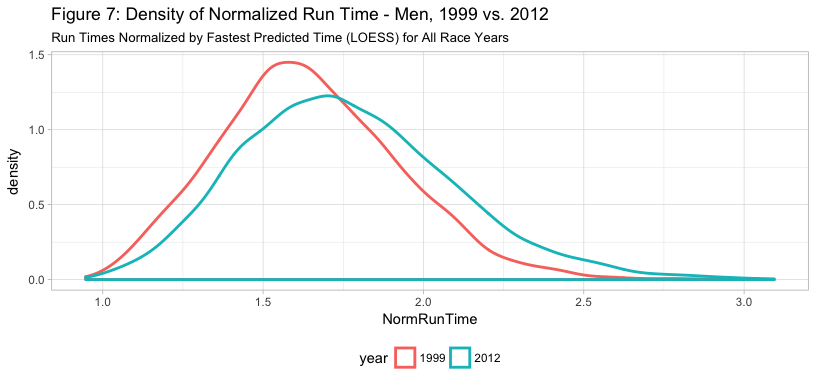
We use the predicted values to normalize the raw 1999 and 2012 race results. The Cherry Blossom Race requires all ten-mile runners to be at least 12 years old and data for individuals older than 80 is extremely sparse. Therefore, only run times for ages greater than 11 and less than 81 are considered. To control for gender-based performance differences, we separate age grading for men and women. We explore the results of implementing age grading using LOESS in the next section.

## Results of Age Grading

LOESS regression predictions allow us to compute age normalized run times for men and women that entered the 1999 and 2012 Cherry Blossom Ten Mile Run races. We first explore individual gender results for both years considered. We then compare the two genders to determine which performed better.

#### Comparing Men’s Run Times for 1999 and 2012

When naïvely comparing median results, runners in 1999 underperform (92.65) relative to runners in 2012 (92.42). This comparison does not account for the age distribution of runners in the field. Given our age grading implementation, we compare both years, accounting for age.



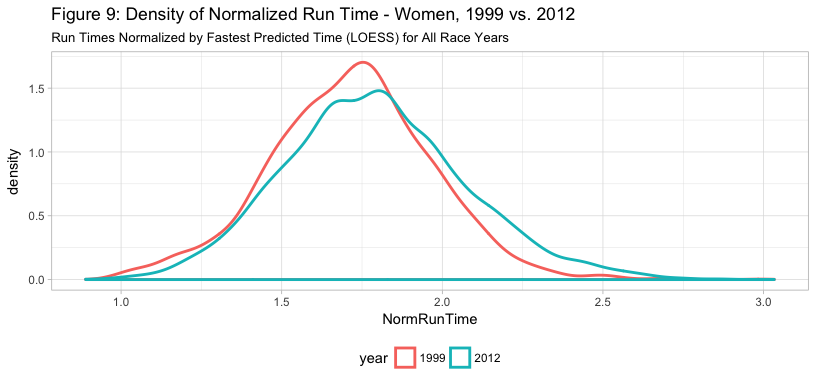
Reviewing the density plot in Figure 7, the 1999 participants age graded run times are shown to be centered just above 1.63 times the fastest runner’s time. 2012 is centered closer to 1.75 times the fastest runner’s time. 2012 also exhibits a wider, flatter curve with a less prominent median. Additional evidence for slowness in 2012 can be clearly seen in Figure 8. The QQ plot shows a pull toward 2012, especially for runners in slower quartiles. These two distributions are obviously not the same and indicate that, when accounting for age, 1999 runners performed better than 2012 runners.

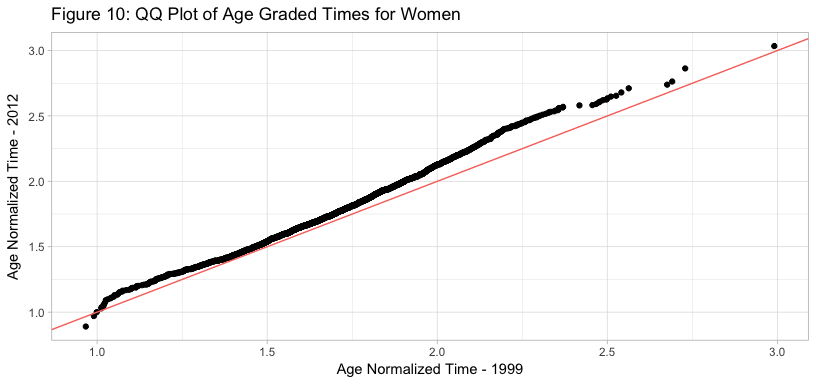
As mentioned in the previous section, the number of runners in the 2012 race has increased significantly. However, we also see a clear shift towards a larger median. This might signal that the growth of the race’s field is mostly in the recreational category of runner- a casual athlete, as well as persons seeking to walk the race route. The second statement seems well supported by the long tail distribution of 2012 runners. In 1999, the race effectively cut-off at the 2.5 times the best runner’s time, while in 2012 the race extended beyond three times the best runner’s time. We could infer that the added participation is largely in mid-pack and lower runners, and not in the faster part of the men’s field.

In fact, male runners in 1999 were found to have a statistically significant difference from male runners in 2012 in the normalized for age run time. Runners in 1999 were a mean of 0.12 faster than in 2012 (p<0.0001). This result differs from the conclusions resulting from naïvely comparing median run times given at the beginning of this section.

#### Comparing Women’s Run Times for 1999 and 2012

We also compare women’s run times between 1999 and 2012 using the same age grading technique. As previously seen in Figure 6, times for women show sharp curvature. Peak run times occur around age 30, and then increase sharply. Of particular note is the amount of run time variation for women in their late 70s. LOESS helps to smooth this variation when using age grading.





Comparing the results of women’s times from 1999 and 2012 reveals less of a difference than we saw for the men. However, 2012 results are very similar to those of the men, with a wider distribution and flatter center of age graded run times. The median run time increases from 1.73 to 1.78, and the mean from 1.72 to 1.80. The interquartile range has also expanded from 0.33 to 0.38.

As seen in the previous section, the increase in female participants expanded dramatically. The distributions of women’s normalized run times for 1999 and 2012 are more similar than the distributions for men. While additional results have been added to the tail of the distribution, the overall length of the distribution has not changed in the same manner that the men’s times have. The maximum race length is still about three times the first finisher for each age group.

Inspecting the quantile-quantile plot provides additional clarity. Similar to the men, results for women in 2012 tend to pull upward from the distribution of 1999 results. This indicates 2012 slightly underperformed 1999 for the women participants. Given these distributions are so close, we also formally compare them with a two-sample *t*-test.

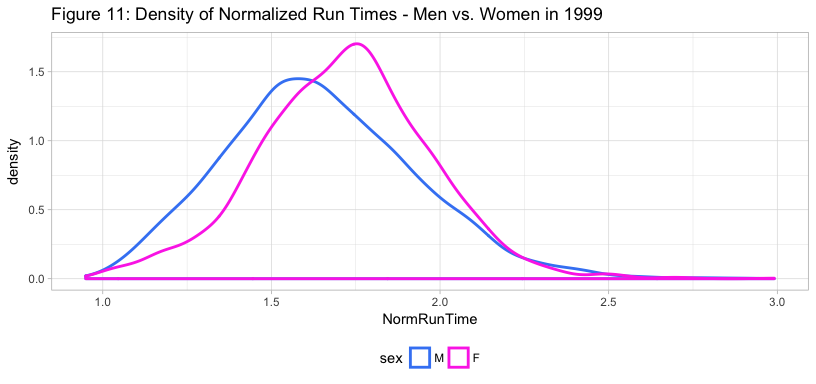
A two-sample *t*-test finds a significant difference in the mean for the two years of women’s race results. While there was a similar normalized median, the difference of normalized mean running times is approximately 0.07 (p<0.0001), indicating the 1999 field was faster given age graded run times. Naïvely comparing run times with no age grading also shows women in 1999 outperform women in 2012.

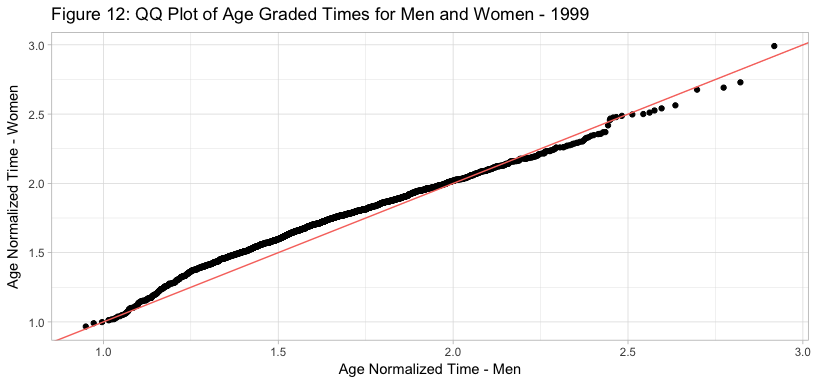
#### Comparing Age Graded Run Times Across Genders

Age graded run times differ when comparing 1999 results to 2012 results for the same gender. In this section we explore gender comparisons directly. We test men’s age graded run times against women’s age graded run times in the Cherry Blossom Ten Mile Run for 1999 and 2012.

##### **Gender Comparison Results for 1999**

Density and quantile-quantile plots are given for the 1999 field of runners in Figures 11 and 12.



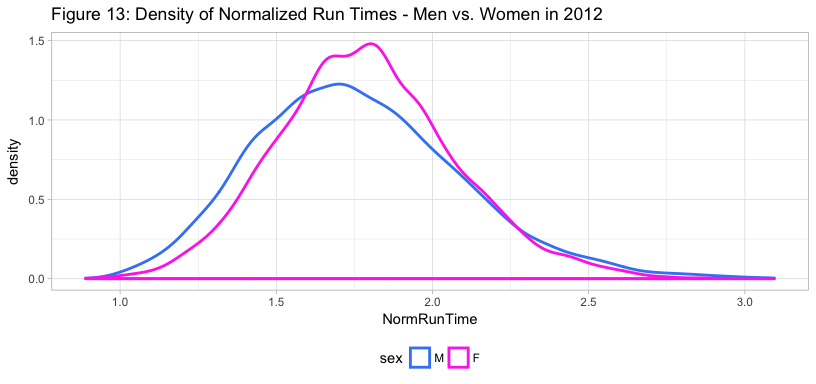


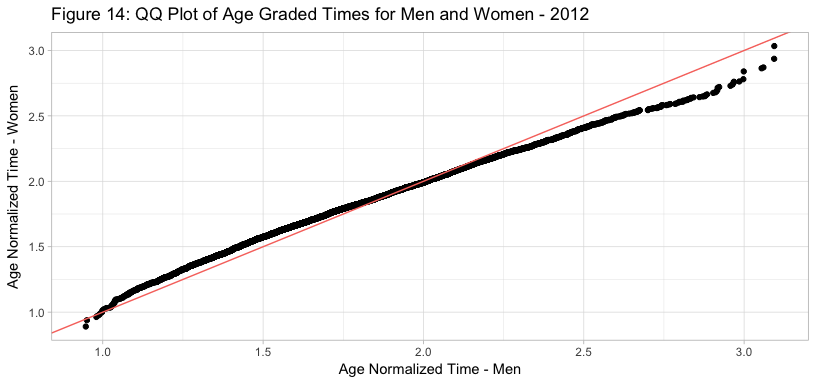
In 1999, we can infer that the performance of the men’s field is more closely aligned to the elite portion of their field than the women’s field is to their elite competitors. The median male performance is 1.63 times the run time of the first-place finisher for their age, while females are above 1.72 times. While men seem to be faster when accounting for age, a stronger right skew can be seen in their density curve in Figure 10. This plays out more apparently when investigating the quantile-quantile plot in Figure 12 of men’s versus women’s 1999 run times. This right skew implies that men had more extreme slow runners than women. However, these slow runners do not make up for the significant performance gap for fast runners seen on the left side of the density plot.

A two-sample student’s *t*-test shows a statistically significant difference of the mean relative performance of men and women. Men are 0.07 faster than women (p<0.0001) when comparing age graded results for 1999.

##### **Gender Comparison Results for 2012**

Density and quantile-quantile plots are given for the 2012 field of runners in Figures 13 and 14.





Reviewing the density curves for 2012, the age graded men’s performance has shifted to be very similar to the women’s. The distribution of the women has a narrower interquartile range than the men. Men’s median age grade performance for 2012 is 1.75 while women’s is 1.79, indicating men performed slightly better when accounting for age in 2012. Similar to 1999, men’s results have a larger right tail. This can be seen in the quantile-quantile plot in Figure 14. The performance gap for the fastest runners between genders is not as large as in 1999. This is represented by the smaller space between the pink and blue lines on the left side of the density plot. However, this gap is still large enough to give men a slight edge in run times when considering age. We formally test differences in mean age graded run times.

A two-sample *t*-test finds a significant difference between men’s and women’s run times. However, the performance gap has narrowed compared to 1999. Men still run a faster time relative to their age and gender, with a mean time of 1.77 times fastest time, while women run a mean time of 1.80 times their age and gender’s fastest time.

## Future Work and Conclusions

The comparison of the results of the Cherry Blossom Ten Mile Race between 1999 and 2012 has led to several interesting insights. As seen for the 1999 versus 2012 men’s results, accounting for age can result in different conclusions when comparing two race fields. In the men’s example, age grading does level the playing field. Overall, results show that 1999 runners typically perform better than 2012 runners. Additionally, men outperform women. However, the gender gap in age graded performance is closing. The women’s field in 2012 performed almost as well as the men.

Before exploring applications, one point of caution must be addressed. The exploration of field size for 1999 and 2012 clearly shows an exponentially increasing level of popularity for competitive running. The 1999 Cherry Blossom Ten Mile Run field was significantly smaller than the 2012 field. With rising popularity, additional considerations need to be investigated to objectively compare running performance. For example, more casual runners and walkers now enter into competitive races. Obviously, this change in demographics could easily bias our earlier comparisons. It is worth investigating a skill level effect as part of age graded analysis to control for potential bias.

The sport of running, and this event specifically, have seen rapid expansion for a 13-year time period, with a substantially higher interest in the event from women. Women’s interest in running as a sport has led to the creation of all-female events like the Disney Princess events. There has also been an expansion of women’s running product lines from many brands, with a shifting strategy towards brands that cater exclusively to female athletes and runners.

The event has also moved toward higher participation by mid-pack athletes and walkers. Race organizers should ensure their race is well organized and still well supported for longer durations. In consideration of these athletes, there may be a need for greater, or different medical support services throughout the course and at the finish line. These athletes are also likely the best audience for many race sponsors.

Application of predictive timing in running and other endurance sports holds numerous applications.

The running community holds the integrity of their sport in highest regard. This has spurred a cottage industry of people who attempt to expose people who may be taking unethical steps to improve their race times. This varies from people not following the selected course (either via a cut through or non-foot transportation), as well as people using surrogates carrying their bib to help qualify for prestigious races like the Boston Marathon. Much of this research has been powered via the published results on race websites, as well as aggregator active.com. The researchers have largely looked for suspicious results on a one-by-one basis, but a LOESS-based model based on age using multiple years’ results could lend new insight to questionable results.

Likewise, race organizers could benefit from deeper insight into their event for practical purposes. Race organizers must first schedule road closures with municipal authorities to ensure a safe course for participants. Understanding the age distribution, and the predicted finish times for their participants, might reduce the chances that a runner is not allowed to finish the race due to slow running times. Additionally, race logistics like start waves are also dependent on expectation of the time required for a specific runner to finish the race. While in most large races this is done on a seeding based on prior finishing times for the same distance, or expected finishing time based on other distances, unknown runners could be added to other groups based on a calculated race finishing time based on age.

Fitness devices and smart phone applications for fitness have become a standard for recreational runners. Predictive pacing could hold a competitive advantage to a device that helps an athlete train smarter.

## 

## References

1. “U.S. Road Race Trends.” Running USA, 23 Mar. 2017, www.runningusa.org/2017-us-road-race-trends.
2. Nolan, D., Temple Lang, D. DATA SCIENCE IN R: A Case Studies Approach to Computational Reasoning and Problem Solving. CRC PRESS, 2017.
3. Anderson, Owen. “Running Science”, HumanKinetics.com, <http://www.humankinetics.com/excerpts/excerpts/genes-and-running-performance>.
4. Tucker, R. “The Anticipatory Regulation of Performance: The Physiological Basis for Pacing Strategies and The Development of a Perception-Based Model for Exercise Performance.”, British Journal of Sports Medicine, June 2009.
5. Tucker R., Noakes TD. “The Physiological Regulation of Pacing Strategy During Exercise: A Critical Review.”, British Journal of Sports Medicine, June 2009.
6. Joyner, M. J., “Modeling: Optimal Marathon Performance on The Basis of Physiological Factors.”, Journal of Applied Physiology, Feb 1991.
7. Lorenz, Daniel S., et al. “What Performance Characteristics Determine Elite Versus Nonelite Athletes in the Same Sport?” Sports Health, SAGE Publications, 5 Nov. 2013, [www.ncbi.nlm.nih.gov/pmc/articles/PMC3806174/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3806174/).
8. “Age Grade Calculator.” Runner's World, 22 Sept. 2017, www.runnersworld.com/tools/age-grade-calculator.