

An Analysis of World Marathon Major Winners

Steve Newns

2018-05-04T21:09:30-05:00

In wake of this year's Boston and London marathons (and their almost polar opposite conditions), there certainly was a lot of press over some historic wins:

- Desi Linden becoming the first American woman to win Boston since 1985
- The “Citizen Runner” Yuki Kawauchi winning his first marathon major (also the first Japanese man in to win Boston in 31 years)
- Eliud Kipchoge further cementing himself as one of the greatest marathoners ever (and in my opinion, *the* greatest), winning his 8th straight marathon, and his 9th out of a total of 10 marathons (It's also important to note this sole 2nd place finish was behind what ended up being a world record time)
- Vivian Cheruiyot took the women's victory in London, taking advantage of the faltering of other elites such as Mary Keitany and Tirunesh Dibaba who battled the heat while gunning for Paula Radcliffe's women's world record

Everything seems to have settled down a bit for the spring marathon season, but I thought it'd be fun to do a cursory analysis of the winners of the six current World Marathon Majors.

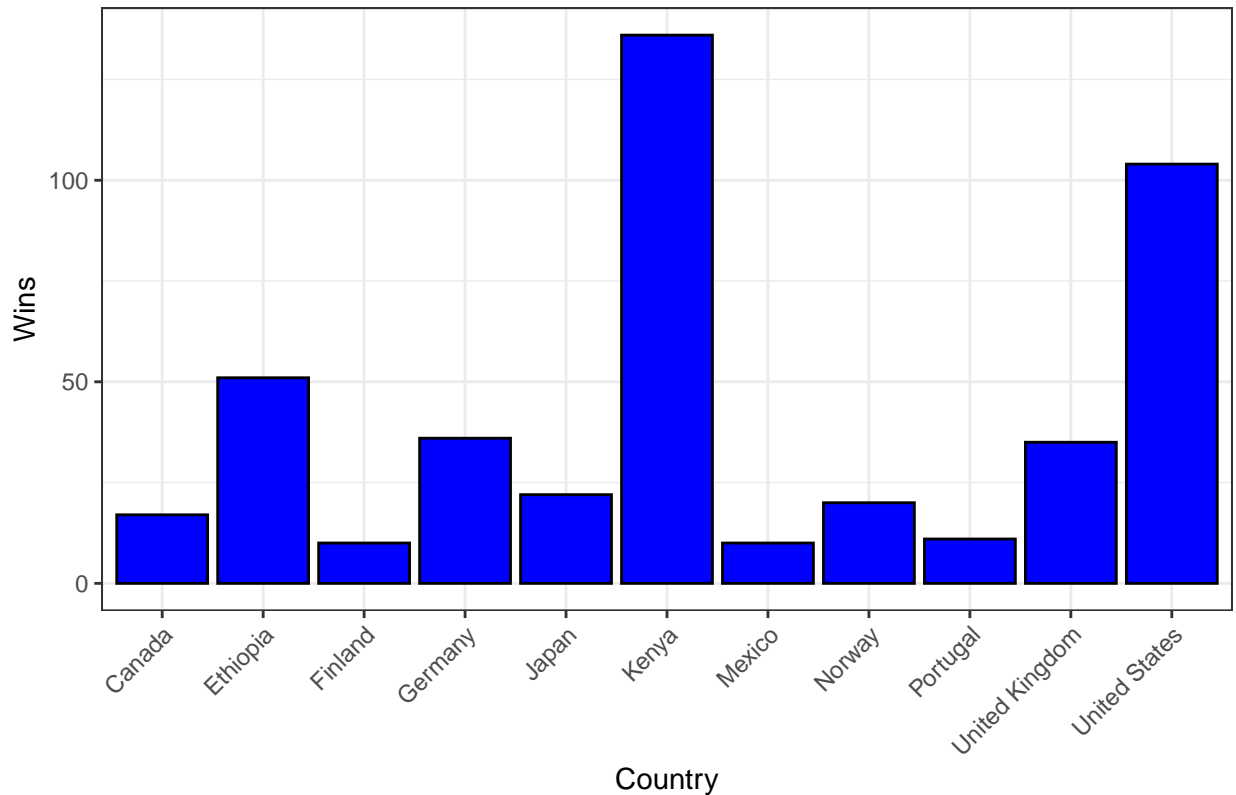
What *are* the World Marathon Majors?

The World Marathon Majors (WMM) is a collection of six annual marathons: Tokyo, New York City, Boston, London, Berlin, and Chicago. And, in designated years, they also include the World Championship (held every 2 years by the IAAF) and the Olympic Marathons. The races are run as a series, and the male and female athletes who score the most points from the qualifying races are crowned the champions of the series (“most points” is defined as the sum of points from a max of two of the qualifying races). For us non-elites, you can qualify as a **Six Star Finisher**, wherein after successfully completing all six of the WMM's, you can receive an official certificate and Six Star Finisher Medal.

Some EDA

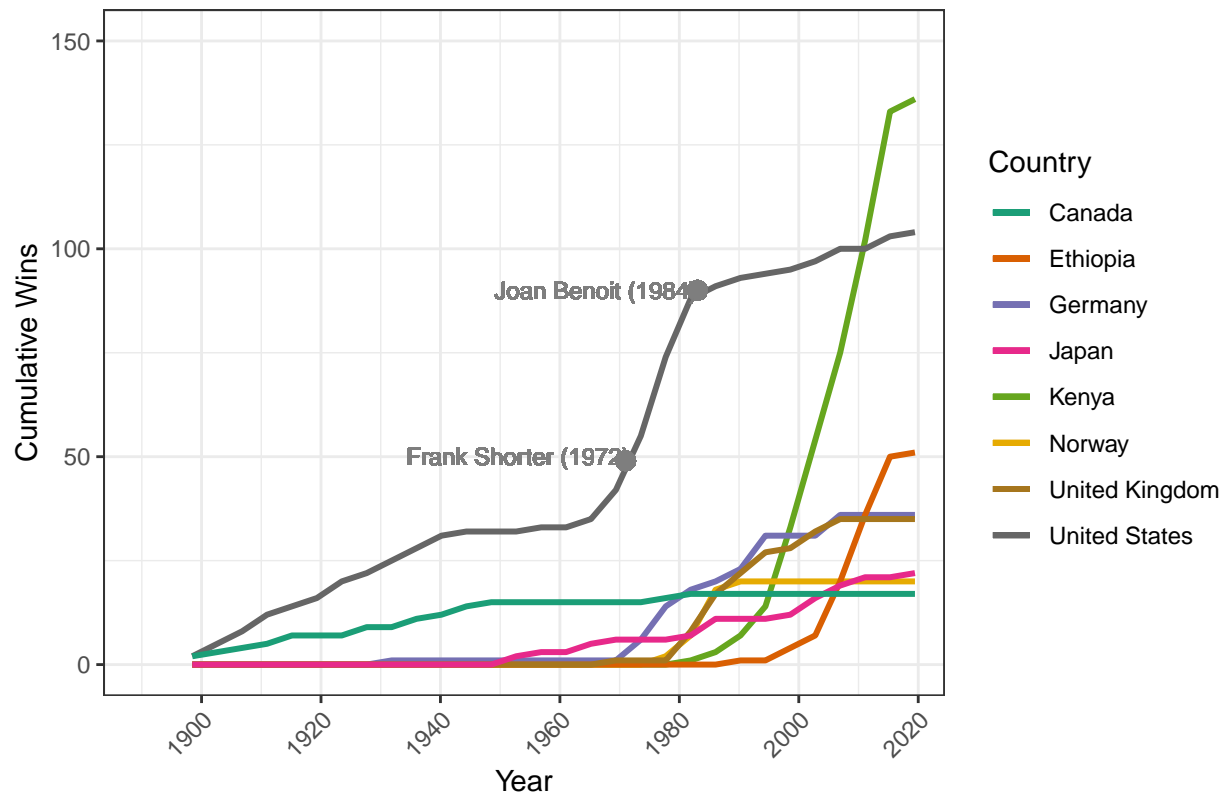
After scraping Wikipedia for the past winners of the six races and cleaning up the data (including having to manually add the female winner for London 2018, Vivian Cheruiyot, since it still wasn't added over a week later, because Wikipedia?), I had a nice data file consisting of, for each race, the marathon, the winner, the year, the winner's gender, and the winner's time. For a quick first plot to look at, I decided on the top 10 countries in terms of WMM wins.

Top 10 Countries in Total World Marathon Major Wins



The general awareness that Eastern Africa (i.e. Kenya and Ethiopia) is dominating long distance running is clearly backed up in this plot, and we see the countries with the next most wins being the US, Germany, the UK, and Japan. It's pretty easy infer a guess as to why, after Kenya and Ethiopia, these specific countries have the most victories. Five of the six major marathons take place in these countries, which I'm guessing likely results in a larger amount of runners from the host country participating, relative to other countries, providing said home countries with more chances of producing the victor. But when did this domination by the East Africans begin?

Cumulative Major Marathon Wins by Country Over Time (1897–2018)

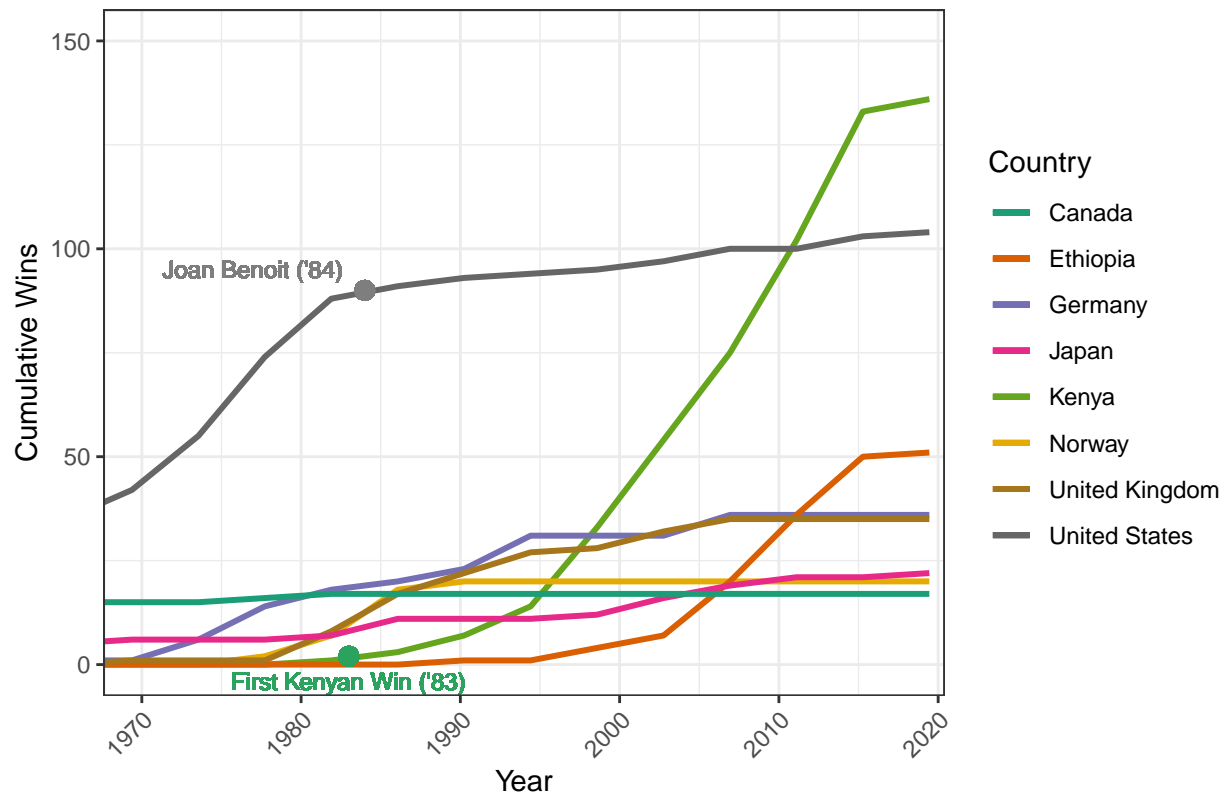


Noting the *incredibly important caveat* that Boston was the only marathon in the current WMM's that existed from its initial running in 1897 until the first running of the New York City Marathon in 1970, I believe it's safe to assume the fact that international travel was not as accessible and common in this time period is why only individuals from countries in close proximity to these races (i.e. US and Canada) saw their counts climb. This makes this chart pretty useless unless we narrow down on the time period that we are looking at.

But, before leaving this chart completely, we can note that German and UK wins started picking up soon after the first couple of their respective marathons, with Berlin starting in 1974 and London in 1981, respectively. The initial plateau for the US and Canada around the 1950's looks to start relatively close to the start of Japan's win count. We can also note that the number of US winners starts to climb exponentially throughout the 1970's, I would assume thanks to the establishment of the NYC and Chicago marathons in this decade, along with the inclusion of women in the Boston marathon. The start of this spike also just barely predates Frank Shorter's gold medal at the 1972 Olympics, which is credited as starting the running boom of the 1970's in the US, which then is said to have culminated with Joan Benoit Samuelson's gold medal in the 1984 Olympics.

But now, onto the limited time period:

Cumulative Major Marathon Wins by Country Over Time (1970–2018)



Here, we see that the first Kenyan win was in 1983, but the country’s rise in marathon victories did not kick-start until the early-to-mid-1990’s. Ethiopia’s win count started a bit later, with their first victory in 1989 and their uptick in victories starting in early 2000’s. We also see some plateaus from the other top countries, with the UK being in a drought since 2008, after 7 wins from Paula Radcliffe over a 6 year span that started in 2002. Germany has just 2 wins in past 20 years, with both of them also occurring in 2008.

Next, I decided to look at who has won the most WMM’s.

winner	country	gender	count
Grete Waitz	Norway	Female	11
Bill Rodgers	United States	Male	8
Ingrid Kristiansen	Norway	Female	8
Clarence DeMar	United States	Male	7
Paula Radcliffe	United Kingdom	Female	7
Uta Pippig	Germany	Female	7
Catherine Ndereba	Kenya	Female	6
Eliud Kipchoge	Kenya	Male	6
Mary Keitany	Kenya	Female	6
Rosa Mota	Portugal	Female	6

We can see that Grete Waitz (the first woman to run the marathon under 2:30:00, in NYC in 1979) is alone at the top of leaderboard, and who, along with fellow dominant Norwegian Ingrid Kristiansen, make up the top two women, and two of the top three runners overall, in terms of WMM wins. It also appears that although Kenya has many WMM wins in aggregate, the total must be quite spread out, as we do not see a Kenyan runner in the top 10 total wins until the tie between Catherine Nderebe, Kipchoge, and Keitany. Kipchoge and Keitany are still actively racing, however, and it’s never a safe bet to count them out of any race. Let’s take the top five runners here and see where their WMM wins occurred.

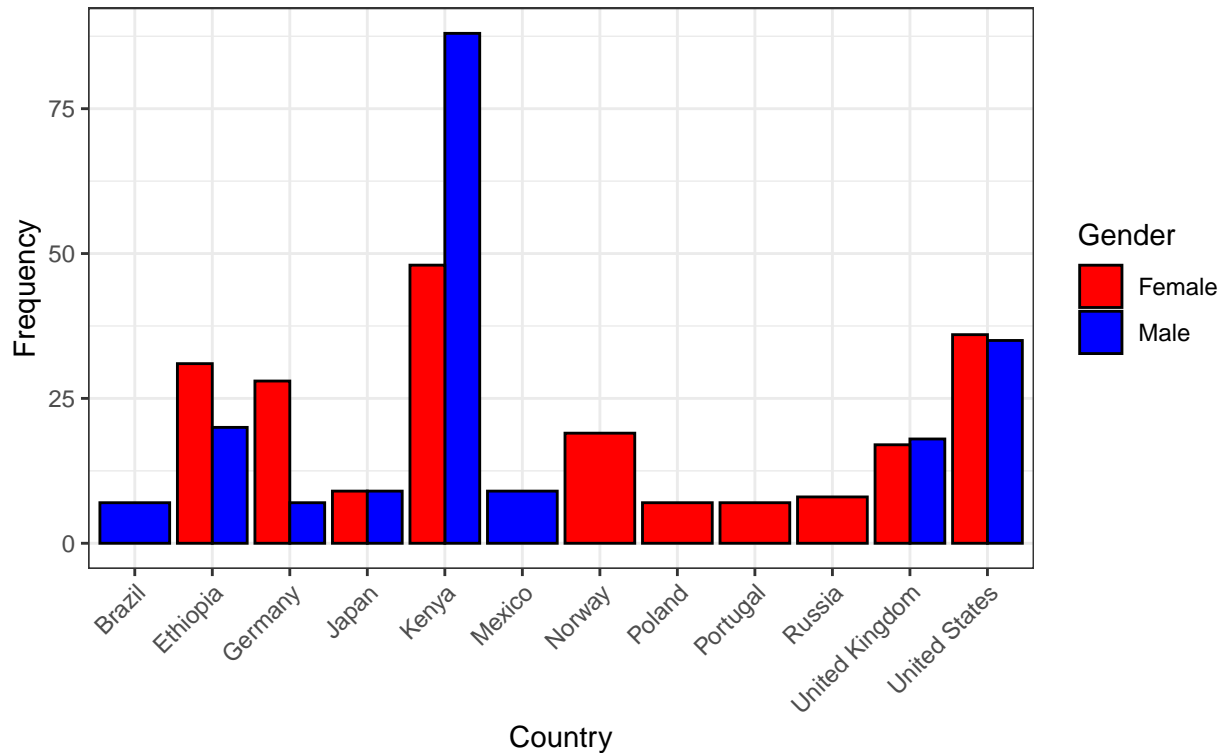
winner	marathon	count
Bill Rodgers	Boston	4
Bill Rodgers	NYC	4
Clarence DeMar	Boston	7
Grete Waitz	London	2
Grete Waitz	NYC	9
Ingrid Kristiansen	Boston	2
Ingrid Kristiansen	Chicago	1
Ingrid Kristiansen	London	4
Ingrid Kristiansen	NYC	1
Paula Radcliffe	Chicago	1
Paula Radcliffe	London	3
Paula Radcliffe	NYC	3
Uta Pippig	Berlin	3
Uta Pippig	Boston	3
Uta Pippig	NYC	1

Waitz was certainly partial to NYC, totaling *nine* wins (out of a total of ten attempts in NYC). The fact that every single marathon major win by Clarence DeMar was in Boston isn't that interesting, as it was the only WMM around during his career in the early 20th century (seven marathon wins is obviously still impressive, nonetheless). Meanwhile, Bill "Boston Billy" Rodgers actually split his WMM wins between NYC and Boston, and Paula Radcliffe split six of her seven wins between her home nation's race (where she also set the current women's world record) and NYC.

Going back to that initial wins-by-country bar chart, we can split it out by gender, filtered down to the years in which both men and women competed and to having more than 5 wins for either gender in order to get a clearer picture of the distribution.

Total Major Marathon Wins by Country by Gender (1966–2018)

***At least 5 wins



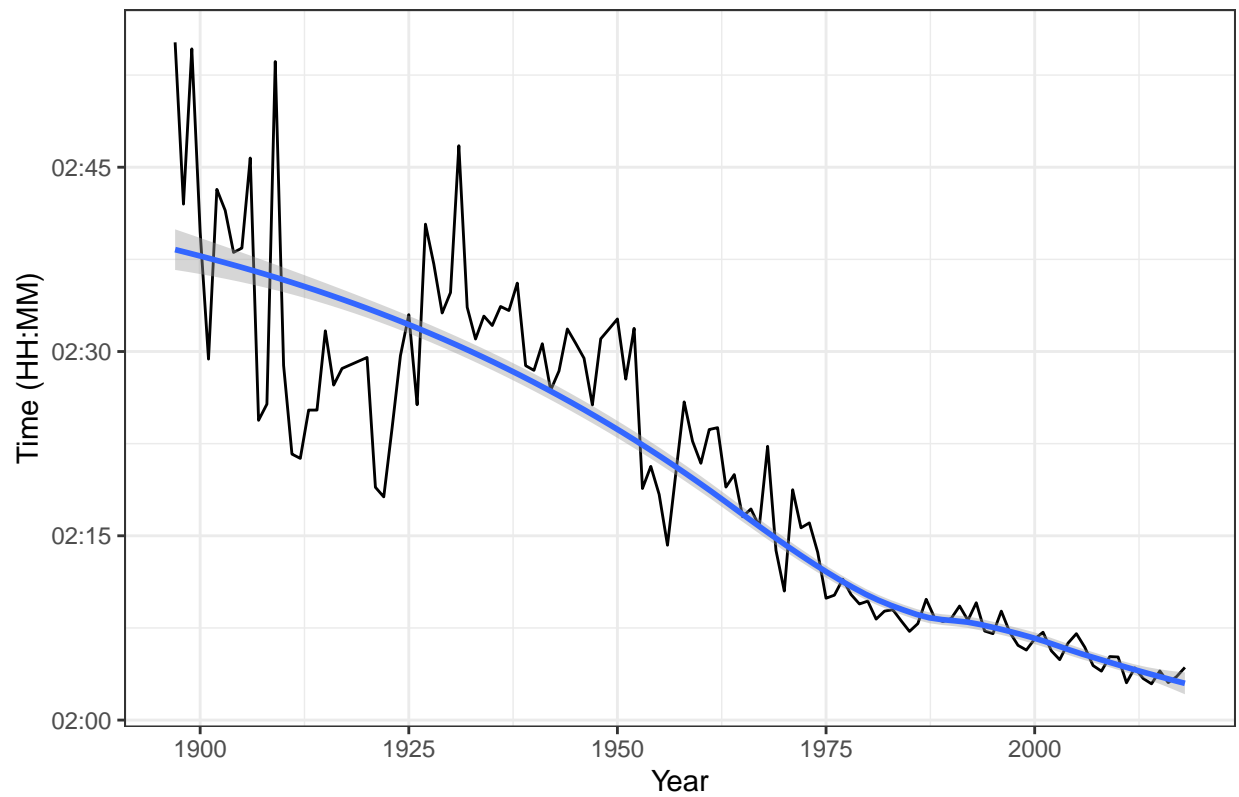
So, Kenya clearly still reigns supreme, having the most male and female winners, while the female runners of Ethiopia and Germany have performed better than their male counterparts. Many countries have had only female winners (see again the dominance of Waitz and Kristiansen in Norway's count and Rosa Mota for Portugal's), while Mexico and Brazil are the only countries with at least five wins that were all males. Then we have Japan, the UK, and the US being evenly split in wins by gender, with a slight favor to the women for the US.

Next, we can look at some visualizations concerning the winning times.

Time

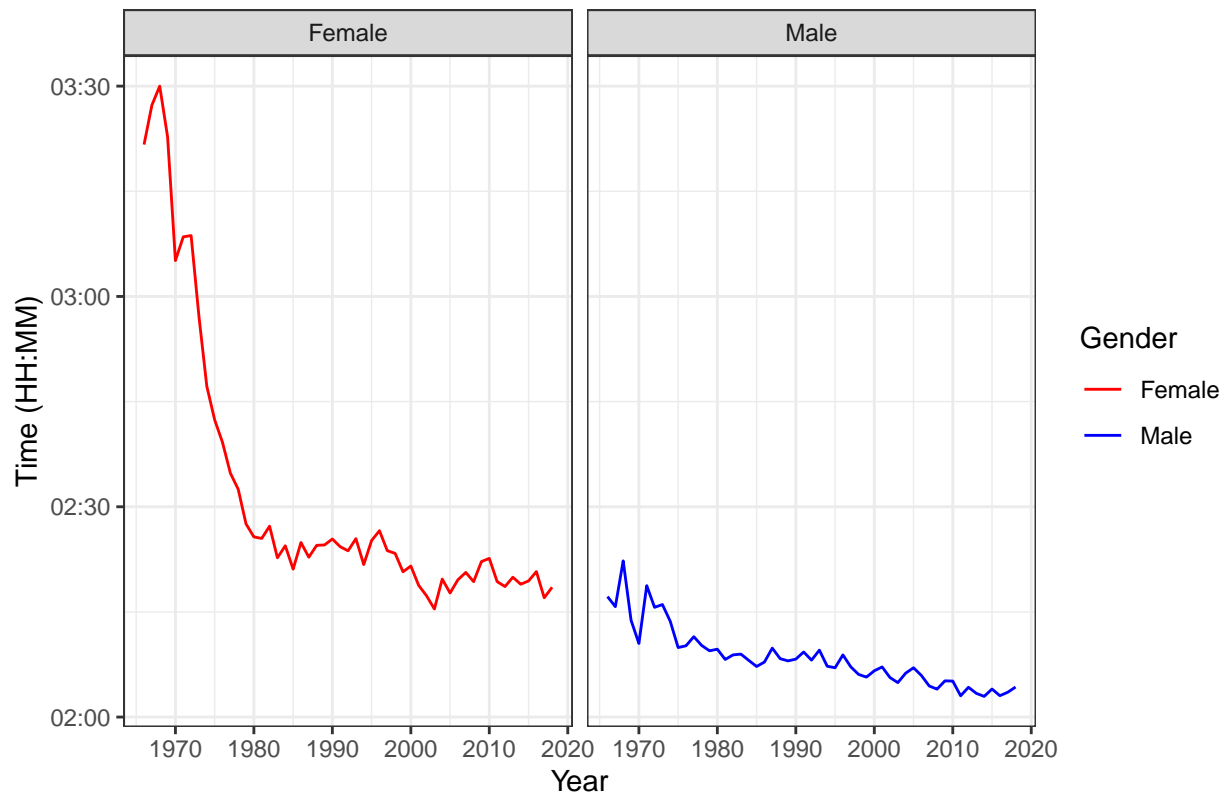
The first thing I wanted to look at was the trend of the fastest winning times per year over the entire time period of the dataset to get an idea of the general trend of winning times over the years (I chose the *fastest* winning times per year because, in a plot of all of the time values, the sudden inclusion of women in the early 1970's produced an incredible wild-looking right-hand side of the plot):

Trend of Fastest Winning WMM Time by Year (1897–2018)



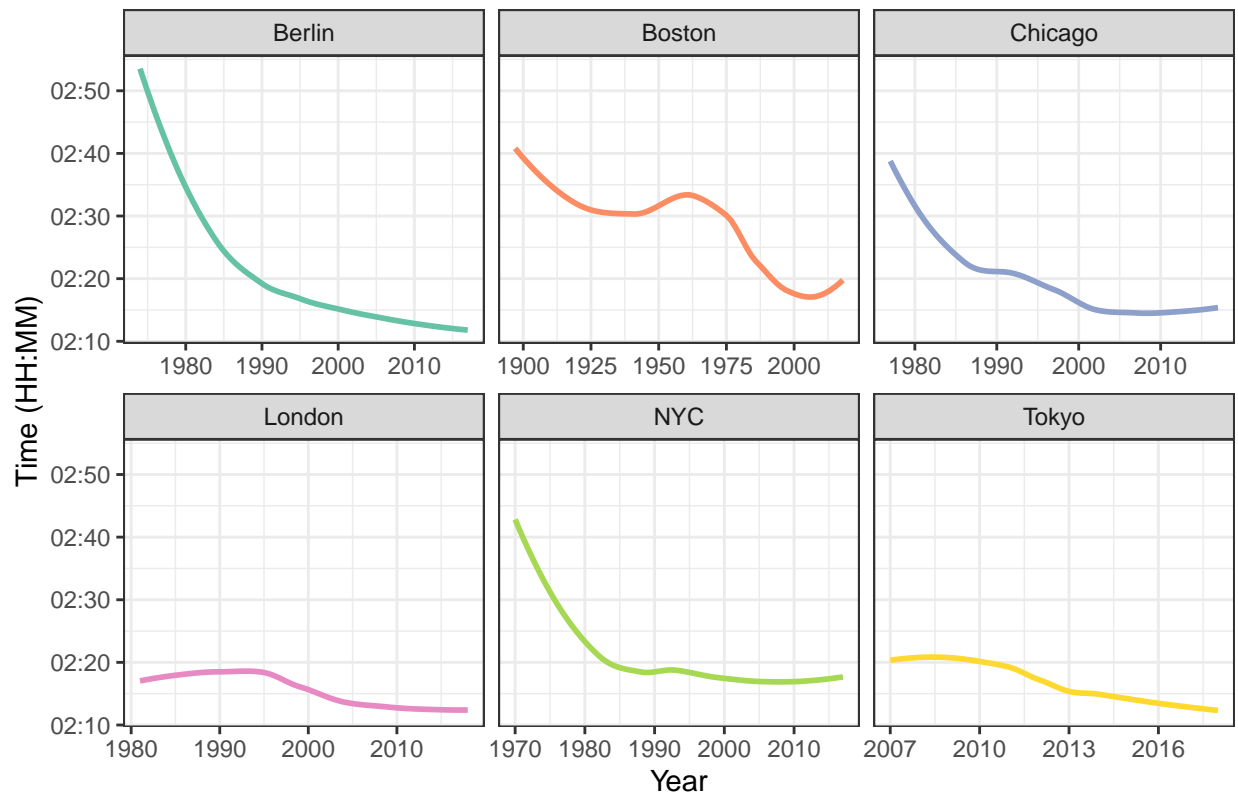
We can see there was a steady, mild decline over the years until the mid-1980's wherein the decrease starts to level off a bit, suggesting that there are only marginal gains to come for the fastest winning WMM times in the future. We can also split out all winning times by gender and limiting to years when both men and women competed.

Trend of Fastest Winning Times by Year by Gender (1966–2018)



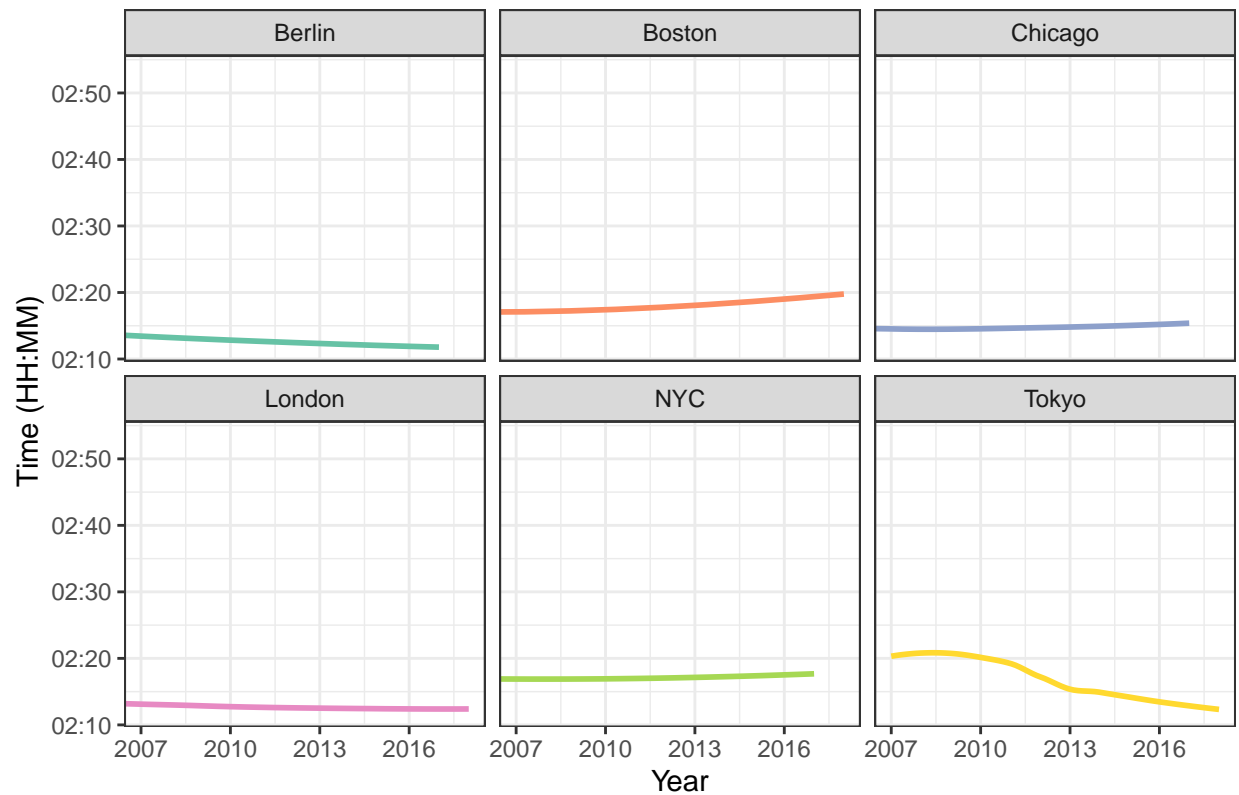
It appears the fastest winning times for women started out quite slow, relative to their current times, followed by a remarkably quick drop between 1970 and 1980. Then, the times start to resemble the slow, declining pattern that is present in the men's plot. These winning times line plots can be split out by marathon as well, where we can then include all winning times, not just the fastest of the year.

Trend of Winning Times by Year by Marathon (1897–2018)



We *could* suggest that the low initial time for the Tokyo marathon here could be due to its relatively recent inception into an era of fast marathons, but the fact that London started about twenty years earlier and is even lower doesn't give that theory too much credence. London winning times also appear to have gotten *slower* in the first fifteen years of the race before starting a steady decline to a now current plateau. We also see an incredible drop in winning times in Berlin, and Chicago and NYC have a similar, but less, severe pattern. Boston has quite the variability, with a noticeable *upticks* in the 60's and in the most recent years. A different look at this data could be to look at these trends solely since Tokyo was introduced:

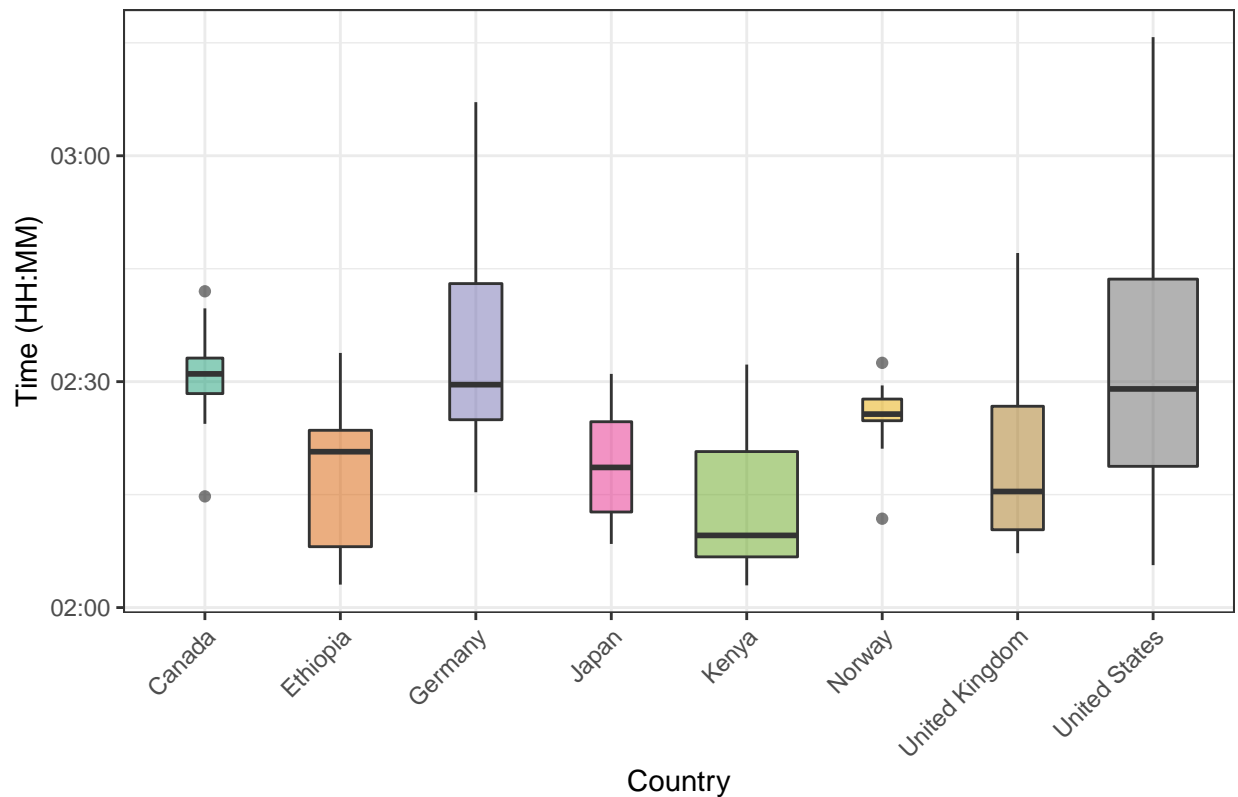
Trend of Winning Times by Year by Marathon (2007–2018)



The majority of winning WMM times have quite a level trend for the past decade, with the exception of the decrease for Tokyo (possibly as more top marathoners started to take part in the new race) and the *increase* in Boston times over the past decade. London and Berlin winning times are *barely* decreasing, while NYC and Chicago have the same trend but in the opposite direction.

We can now break out all winning times out by winning country in boxplots to see how those distributions come out.

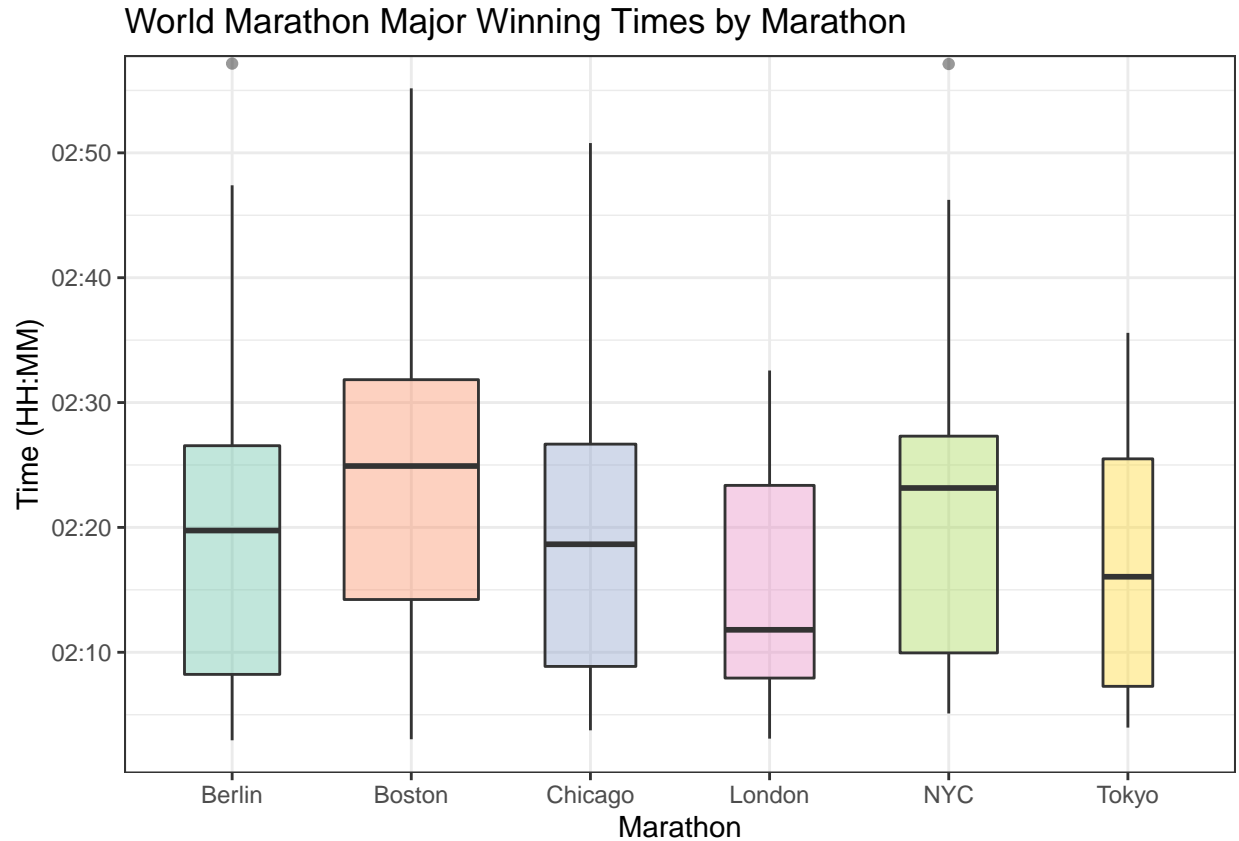
World Marathon Major Winning Times by Country



I'd say the larger relative variability in the US box is understandable, when accounting for the fact that many of these winners won during the early years of the long history of Boston before marathon times started to significantly drop. Germany's wide variance is something to note, for me at least. It was my understanding that runners go to Chicago or Berlin is going for a fast time, as they are the flatter, and therefore one would think faster, courses. I would think more Germans would've been running in Berlin, and therefore would've had quicker times with not too much variability. But, maybe the weather could be a factor here, and it could be something to look into.

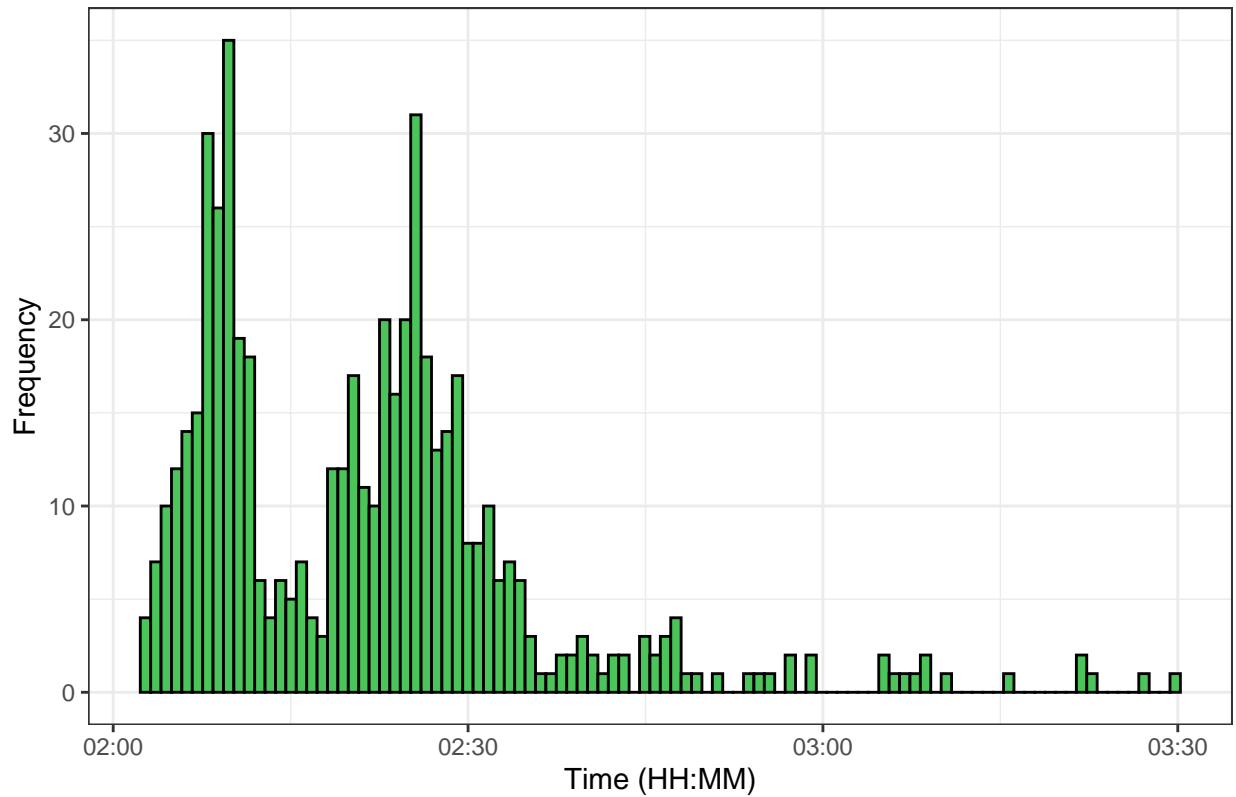
We can see Norway and Canada winners both have very little variability in winning times, and coincidentally both countries have one slow and one fast outlier, albeit with lower sample sizes than Kenya, Ethiopia, and the US. Norway's distribution also appears to be a bit skewed, with its median winning time being located more towards the faster quartile. Also, again, Kenya continues to perform better than other countries, having the fastest median win time (again, with a possible skew). Then, while the UK and Japan both seem to have faster winning times than Ethiopia if going by the median time, Ethiopia, in fact, wins more races. This isn't so surprising since many major marathons are run with the intent of executing some tactical strategy in order to win the race and not so much with a focus on having a very fast time.

Speaking of which, let's look at the times with respect to the races themselves.



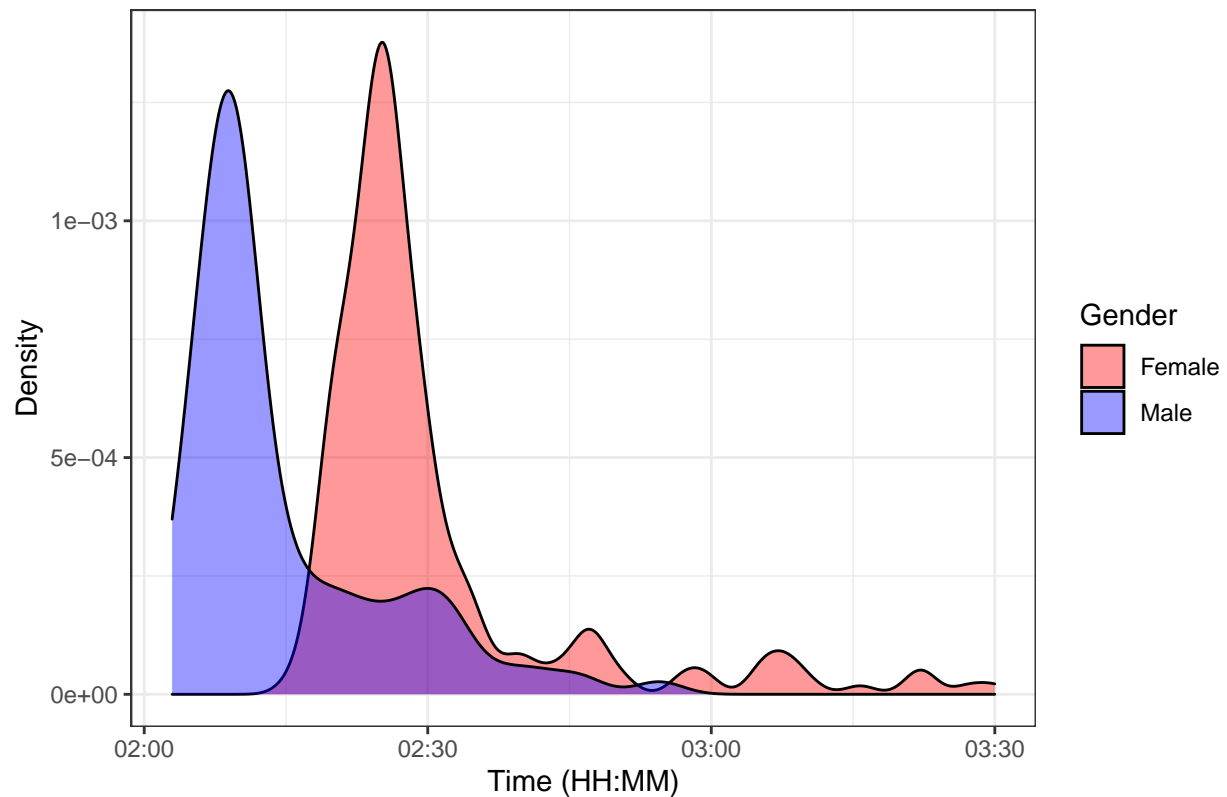
So, Boston and NYC appear to be the slowest of the WMM's, with higher (i.e. slower) quartiles and higher median times. This makes sense to me, as they're both known for their tough hills. The "fast" marathons in Chicago and Berlin seem almost identical to me, with Berlin have a couple more of the slower outliers. But, it looks like, rather than Berlin and Chicago, *London* is actually the race to run if one is looking for a faster time, having the lowest (i.e. fastest) median time, and the fact that this median skews towards the faster quartile. Overall, the variability among marathons seems quite even, as we can see from the heights of all boxes sans Boston being constrained between 2:10 and 2:30. From this plot, Tokyo also looks to have quite a normal distribution, but that remains to be confirmed via a histogram or density plot, which we can move onto now.

Distribution of WMM Winning Times



The bimodal nature of the histogram is apparant, and is almost certainly due to differences in times by gender. We can see that one peak of winning times centered in the mid 2:20's, most likely for the women, and the other peak centered in the lower 2:10's, most likely for the men.

Distribution of WMM Winning Times by Gender



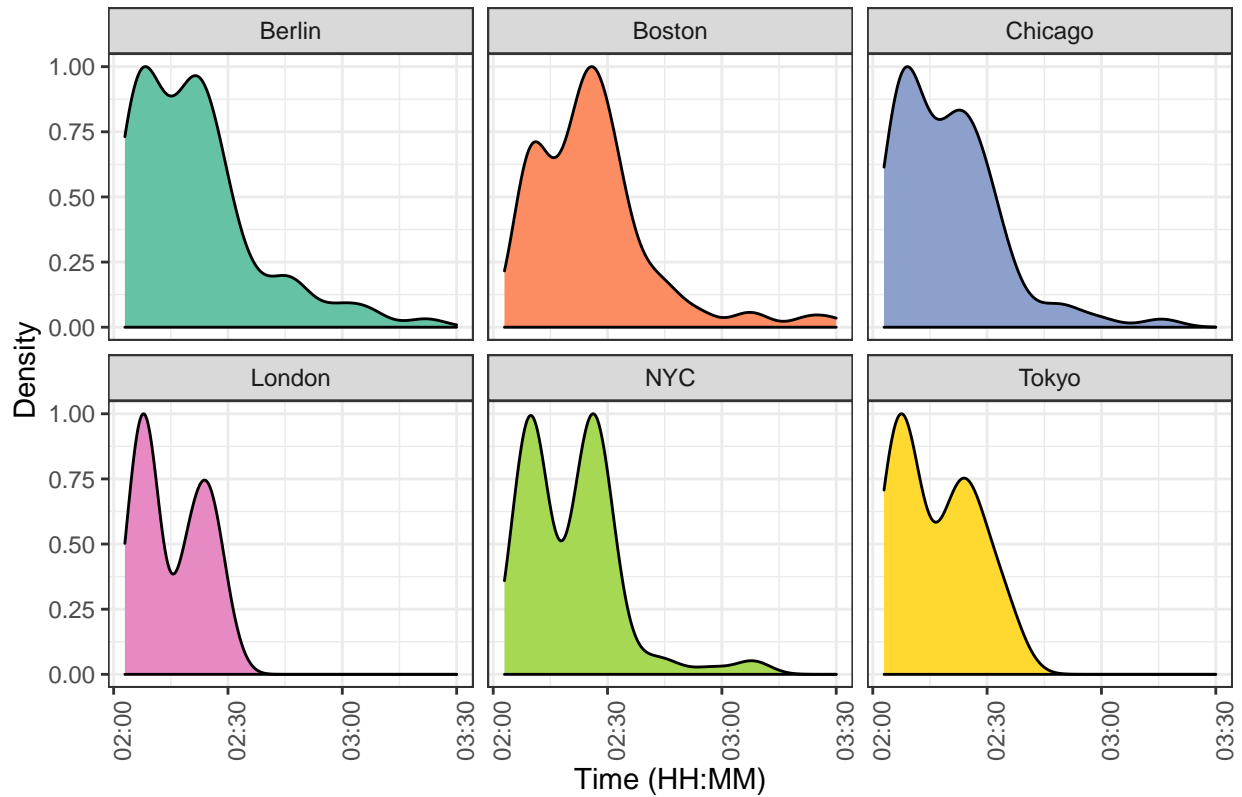
Indeed, these densities confirm my (pretty obvious) assumption. They also further highlight the right-skew of both distributions. Other than that, it appears that the spread of both density peaks is around 15 minutes in either direction from their peak. This can easily be inspected further:

gender	medianTime	stdDevTime
Female	02:25:56	28481
Male	02:10:11	27700

Okay, so the peak of the women's time distribution was in the mid 2:20's and the men's was actually right at 2:10, and the distributions had remarkably close standard deviations of about 19 minutes. We can split these density plots out by marathon, both overall and also then by gender.

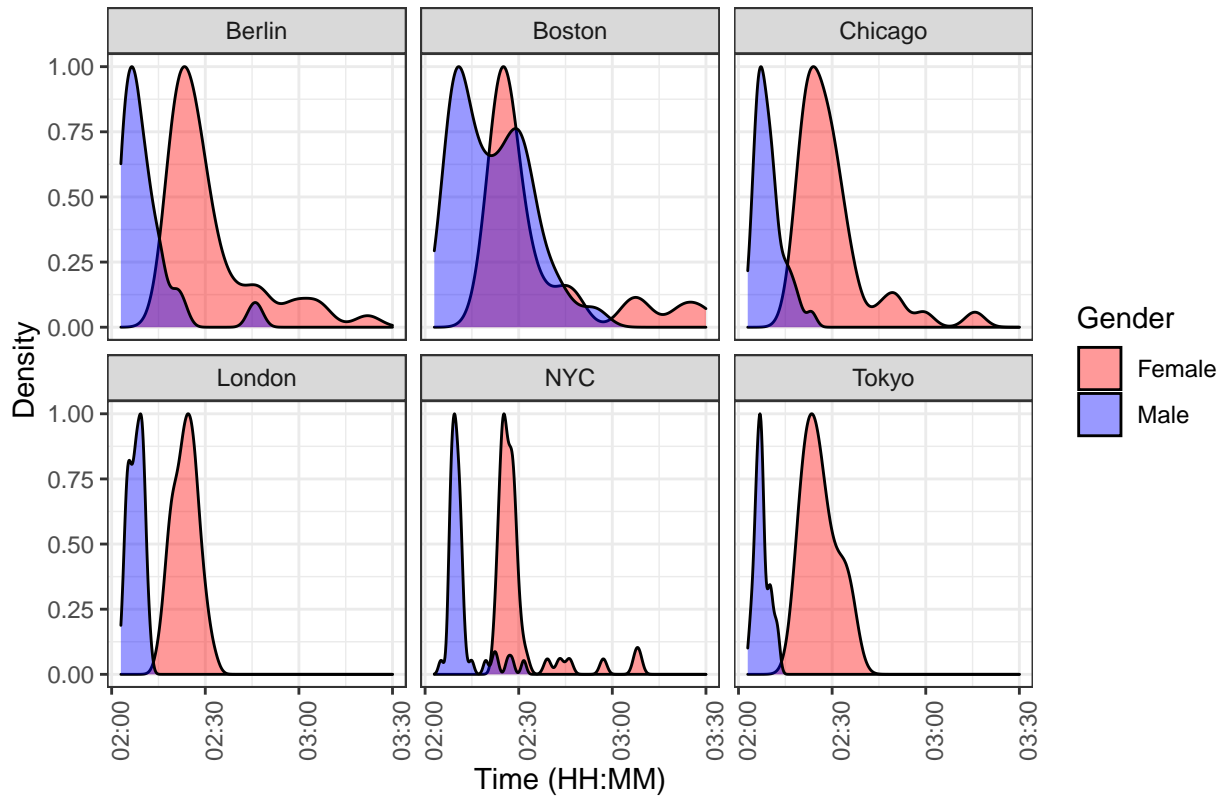
First, just by marathon:

Density of Winning Times by Marathon



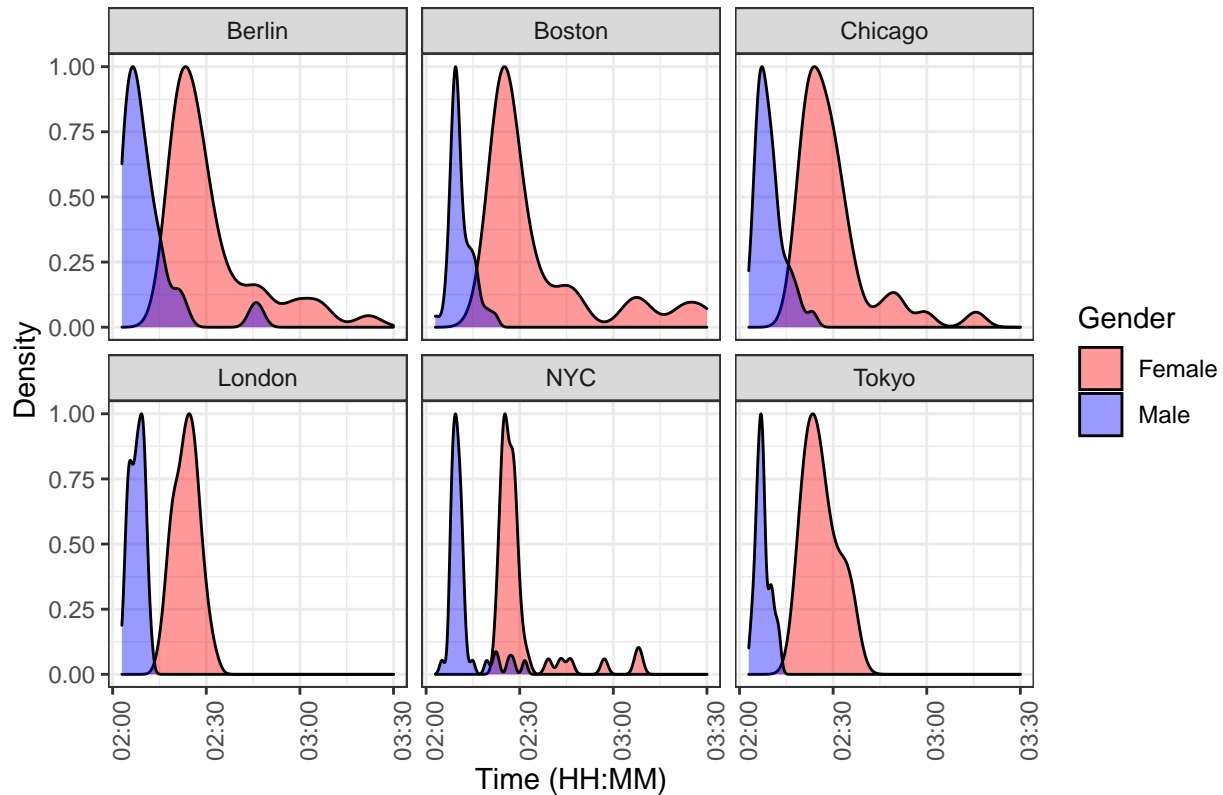
For some countries, the bimodal nature is more pronounced, such as in London and NYC, with NYC's peaks looking quite similar, and overall having a noticeable right-tail, relatively speaking. Berlin and Chicago have less pronounced peaks, and we can see London and Tokyo have much taller densities in the lower-two hour range, in contrast to Boston, who is the only marathon wherein the slower mode has a noticeably taller density in comparison to the faster mode. This info can also be split out by gender.

Densities of Winning Times by Marathon by Gender (1897–2018)



It seems that Boston's long history has caused a bimodal distribution to develop even just for the male winning times, and not solely for overall times. Also, recalling the fact that the boxplots for London and Tokyo were the 2 marathons with no outliers, we can see that they are the only two density plot pairs here without a pronounced right tail. It appears that NYC has two very non-variable (skinny) modes, along with multiple very small bumps going down the x-axis to the right, suggesting outliers there occur more sparsely. The males in Chicago, Berlin, London, and Tokyo all have skinnier densities as well, indicating less variety in winning times for males in these marathons in comparison to the females, whose densities are quite wider. For another picture, we can, again, filter down to just years where both men and women competed.

Densities of Winning Times by Marathon by Gender (1966–2018)



NYC still the least variable distributions for both genders, again followed by London. NYC, Chicago, Boston, and Berlin (and arguably Japan) still have that right tail for the women's distribution while Tokyo and London still do not. The tight densities for most male distributions suggest a consistent range of winning times at these marathons, as well as for the women in NYC, relatively speaking. Berlin, Chicago, and Boston also have quite similar pairs of density plots.

Conclusion

Overall, this was just a quick analysis I thought to go through in light of all the recent excitement of London and Boston this year. I'm pretty sure most of these inferences won't change much after the rest of the WMM's for 2018 are ran, but I'll probably still update the dataset and these plots all the same. I may also add the Olympic and World Championship races as well.

The code for the scraping and cleaning can be found [here](#) and the code for this writeup can be found [here](#). If you've got any comments, feel free to let me know on [Twitter](#), [GitHub](#), or [LinkedIn](#).