How to train like a pro for the Boston Marathon?

W200 - Project 2 - Aug 2019

Cecily Sun, Justin Hsia, Nobu Yamaguchi

Purpose:

In this analysis, we will mainly answer the question: What are the factors that have impact on the performance of marathon runners? The goal of the project is to generate meaningful insights and actionable recommendations for future Marathon runners in order to help them train well and achieve more.

The project will include 3 modules of analysis: Exploratory Data Analysis, Correlation Analysis, and Cluster Analysis.

Questions:

Our team raised a couple of hypotheses on the potential factors that have impact on the performance of marathon runners, both internally and externally. Here are the 3 main questions we will answer -

- How does weather condition affect the overall finishing time for Boston Marathon?
- How does the elevation change affect runners' pace?
- What's the good pacing strategy for Boston Marathon?

Dataset:

All the raw data are in csv format. Please note that the data for 2013 is missing due to the bombing terrorist attack. Also, the data before 2013 don't have the timing records at each distance(5k, 10k, 15k, etc.). The data in 2014 is missing the record at 15k.

- Boston Marathon Raw Data 2001-2017
- Boston Marathon Elevation Data
- Boston Marathon Weather Data

Link to the dataset:
https://drive.google.com/drive/folders/178ysWRP6XB71MMp9SrCKBwyChlTRk-Wz?usp=sharing

Part 1: Data Cleaning & Manipulation

The first step in our analysis is to do data cleaning and manipulation of over 16 individual csv files. Since data before and after 2014 came from different sources, it becomes more challenging to align data from different sources and combine them into a single dataset for the following analysis. The naming conventions were different, the data types were different, the data representations were different. Especially when it comes to the records of finishing time, some datasets recorded it in H:MM:SS, some datasets recorded in minutes. We had to convert the data into the same format before proceeding to the analysis.

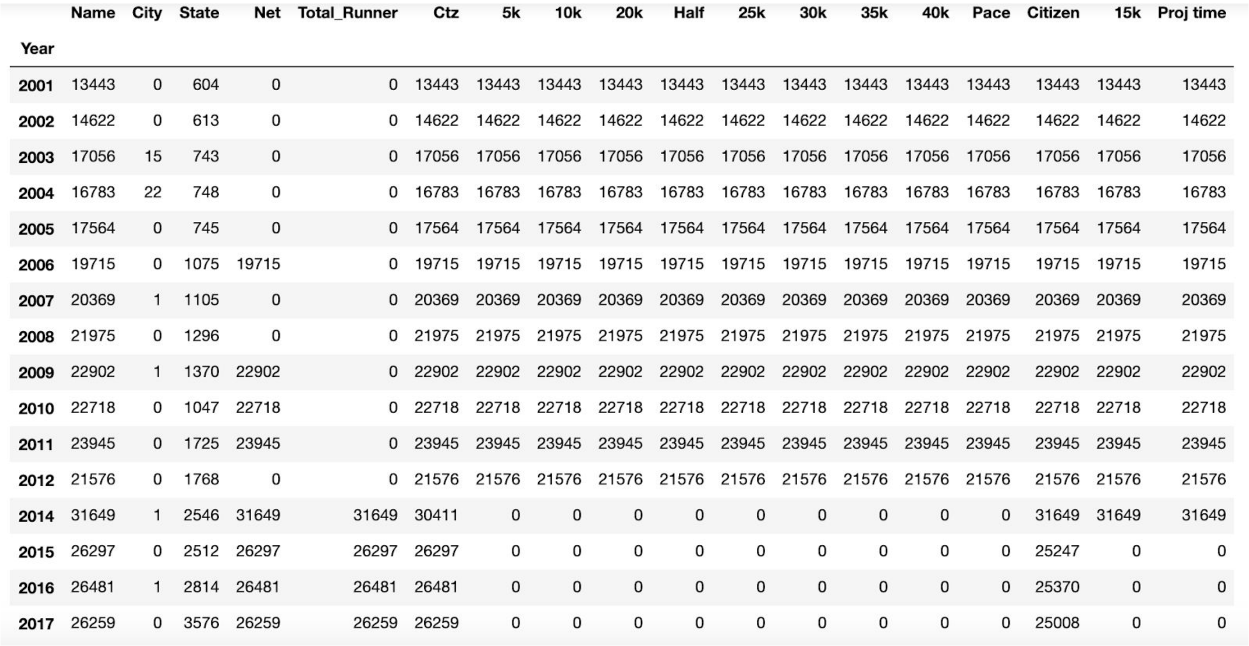
After putting all the 16 years worth of data together, we get to a full dataset with 343,430 rows in total.

Part 2: Sanity Check

Missing Data

The first thing we did in sanity check is to look out for missing values. As is turned out, there were 76 rows missing the most important variable - Official, which is the official finishing time for each runner. Since the number of missings was very small, we just simply dropped them.

Then, we took a look at all the other columns and checked if they have missing values. Below is the table output for this step - The Name column represents the number of rows broken down by year, and the rest of the columns are the number of missing values compared to the total number of rows broken down by year.



	Name	City	State	Net	Total_Runner	Ctz	5k	10k	20k	Half	25k	30k	35k	40k	Pace	Citizen	15k	Proj time
Year																		
2001	13443	0	604	0	0	13443	13443	13443	13443	13443	13443	13443	13443	13443	13443	13443	13443	13443
2002	14622	0	613	0	0	14622	14622	14622	14622	14622	14622	14622	14622	14622	14622	14622	14622	14622
2003	17056	15	743	0	0	17056	17056	17056	17056	17056	17056	17056	17056	17056	17056	17056	17056	17056
2004	16783	22	748	0	0	16783	16783	16783	16783	16783	16783	16783	16783	16783	16783	16783	16783	16783
2005	17564	0	745	0	0	17564	17564	17564	17564	17564	17564	17564	17564	17564	17564	17564	17564	17564
2006	19715	0	1075	19715	0	19715	19715	19715	19715	19715	19715	19715	19715	19715	19715	19715	19715	19715
2007	20369	1	1105	0	0	20369	20369	20369	20369	20369	20369	20369	20369	20369	20369	20369	20369	20369
2008	21975	0	1296	0	0	21975	21975	21975	21975	21975	21975	21975	21975	21975	21975	21975	21975	21975
2009	22902	1	1370	22902	0	22902	22902	22902	22902	22902	22902	22902	22902	22902	22902	22902	22902	22902
2010	22718	0	1047	22718	0	22718	22718	22718	22718	22718	22718	22718	22718	22718	22718	22718	22718	22718
2011	23945	0	1725	23945	0	23945	23945	23945	23945	23945	23945	23945	23945	23945	23945	23945	23945	23945
2012	21576	0	1768	0	0	21576	21576	21576	21576	21576	21576	21576	21576	21576	21576	21576	21576	21576
2014	31649	1	2546	31649	31649	30411	0	0	0	0	0	0	0	0	0	31649	31649	31649
2015	26297	0	2512	26297	26297	26297	0	0	0	0	0	0	0	0	0	25247	0	0
2016	26481	1	2814	26481	26481	26481	0	0	0	0	0	0	0	0	0	25370	0	0
2017	26259	0	3576	26259	26259	26259	0	0	0	0	0	0	0	0	0	25008	0	0

From the table above, we noticed that some geographic information, such as City and State, were not complete. But the number of missing values from these 2 variables are a lot, so we just ignored them.

The majority of Citizen columns are missing values, which won't add value to our following analysis, so we just dropped this column.

For columns Net and Proj time, there are a lot of missing values. We also dropped the two columns as we won't be using these two columns in the future analysis.

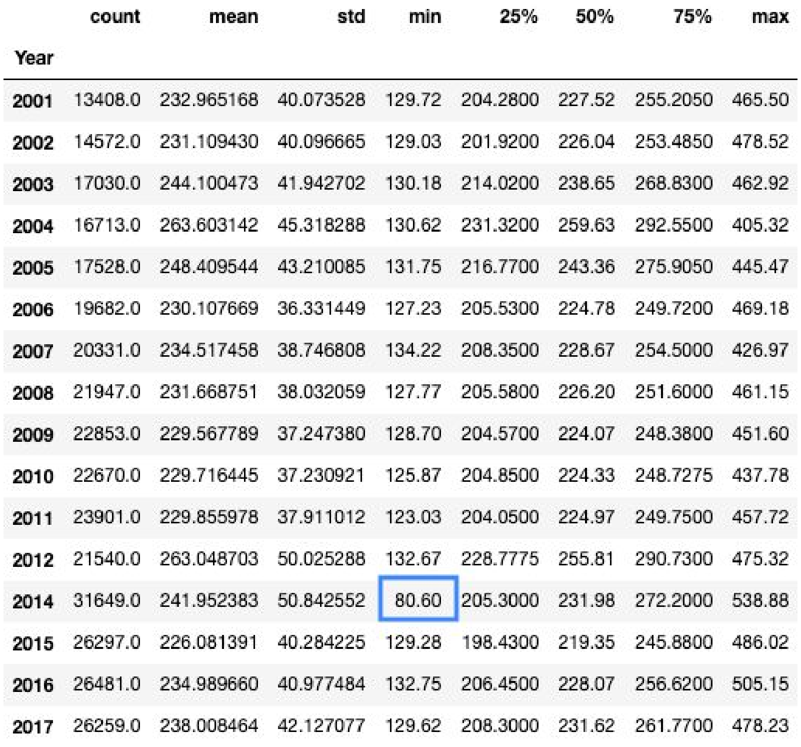
Finally, the distance time data before 2014 were missing due to incomplete data collection. We will ignore the missing before 2014 and only use data after 2014 to perform analysis that require these columns.

Numeric Variables & Categorical Variables

Before moving on to the analysis, We want to do another sanity check on both categorical variables and numeric variables. For categorical variables, we want to check the frequency count and see if it makes sense; for numeric variables, we want to check the min, max, average values to make sure that there won't be any outliers. Sometimes even if there are no missing values in the data, this step is essential to identify the noises in the dataset.

For example, we checked the column Total_Runner by year and compared with with the total number of rows by year. The reason I'm doing this is that the Total_Runner column should be able to match with the total number of rows by year. If not, there must be something wrong here. As it turned out, there were some noises that certain records are from wheelchair racers. So we just dropped these noises as they are a small amount of data, and they won't affect the analysis we will be doing later.

For numeric variable, we mainly checked the most important variable - Official, which is the official finishing time for the runners. Here's a table describing this variable (converted in minutes).



	count	mean	std	min	25%	50%	75%	max
Year								
2001	13408.0	232.965168	40.073528	129.72	204.2800	227.52	255.2050	465.50
2002	14572.0	231.109430	40.096665	129.03	201.9200	226.04	253.4850	478.52
2003	17030.0	244.100473	41.942702	130.18	214.0200	238.65	268.8300	462.92
2004	16713.0	263.603142	45.318288	130.62	231.3200	259.63	292.5500	405.32
2005	17528.0	248.409544	43.210085	131.75	216.7700	243.36	275.9050	445.47
2006	19682.0	230.107669	36.331449	127.23	205.5300	224.78	249.7200	469.18
2007	20331.0	234.517458	38.746808	134.22	208.3500	228.67	254.5000	426.97
2008	21947.0	231.668751	38.032059	127.77	205.5800	226.20	251.6000	461.15
2009	22853.0	229.567789	37.247380	128.70	204.5700	224.07	248.3800	451.60
2010	22670.0	229.716445	37.230921	125.87	204.8500	224.33	248.7275	437.78
2011	23901.0	229.855978	37.911012	123.03	204.0500	224.97	249.7500	457.72
2012	21540.0	263.048703	50.025288	132.67	228.7775	255.81	290.7300	475.32
2014	31649.0	241.952383	50.842552	80.60	205.3000	231.98	272.2000	538.88
2015	26297.0	226.081391	40.284225	129.28	198.4300	219.35	245.8800	486.02
2016	26481.0	234.989660	40.977484	132.75	206.4500	228.07	256.6200	505.15
2017	26259.0	238.008464	42.127077	129.62	208.3000	231.62	261.7700	478.23

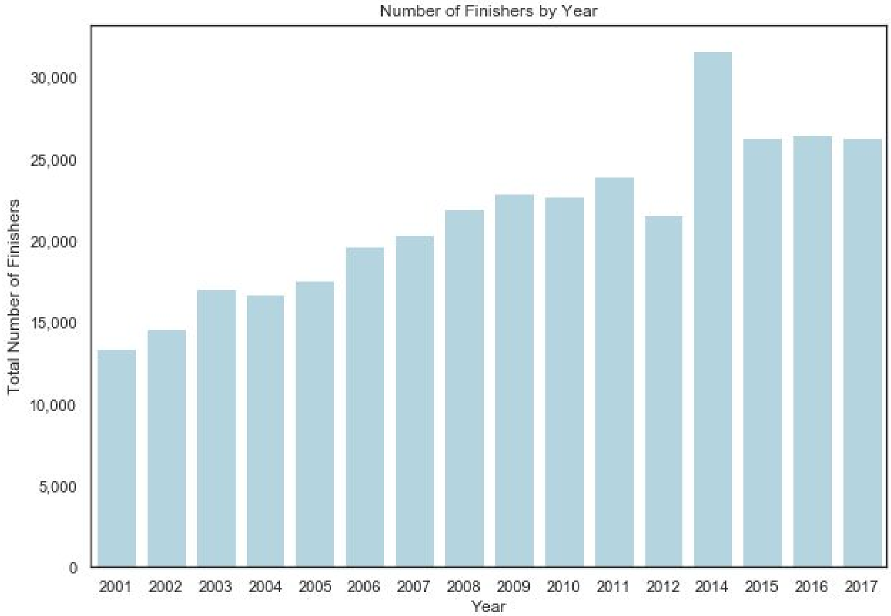
Looking at the table above, we found the minimum finishing time for 2014 looked odd to me. My assumption here is that there might be some records that are not within the main category of the Boston Marathon - such as hand cyclists. After cross-checking with the official 1st finisher data from Wikipedia, we decided to align these records with the official data pulled from Wikipedia. Later on we figured out that the wheelchair racers were labeled with ‘W’ in front of their BIB numbers, which made it easier to pick them out from the main dataframe for the following analysis.

Part 3: Exploratory Data Analysis

In the first part of the project, we will be looking at a 16-year panel data on historical records of Boston Marathon runners. Exploratory analysis helps us to get some insights from the dataset. We want to be aware of such trend in the data that may have impact over the correlation analysis. We want to understand the following questions -

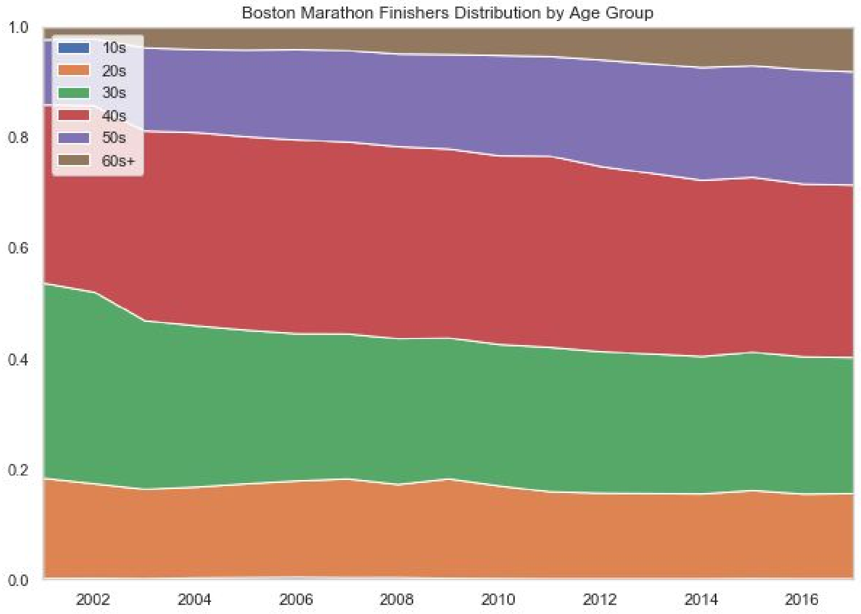
- What’s the distribution of performance(finishing time) look like among all the runners?
- What’s the ratio between male runners and female runners? Is there a shift towards a certain group over the years? The same question can be applied to different age groups.

Total number of finishers by year



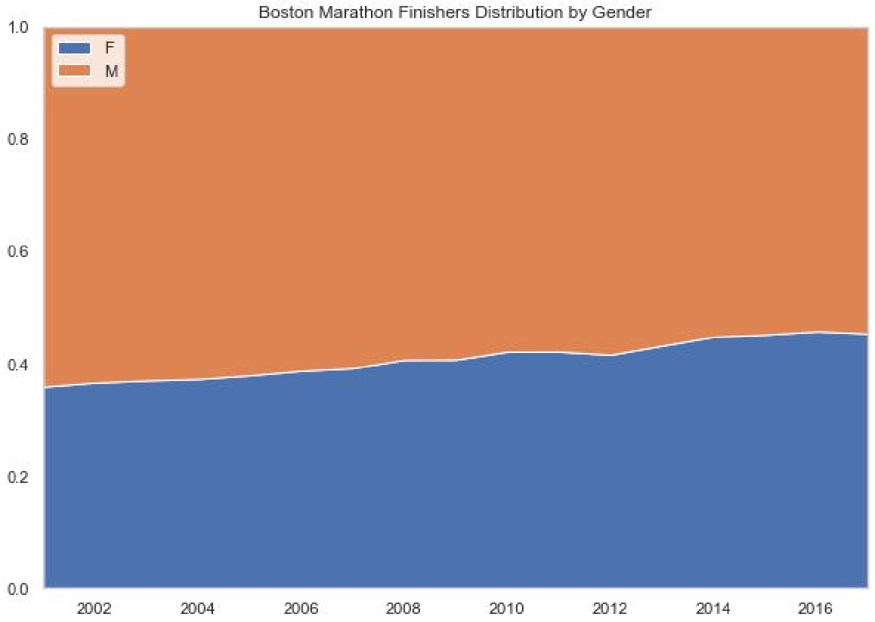
Looking at the chart above, it seems that 2014 has an extremely high number of participants. we double-checked with the official record from Boston Marathon’s website and confirmed that there were unprecedented 32k finishers in 2014.

Total number of finishers by age group



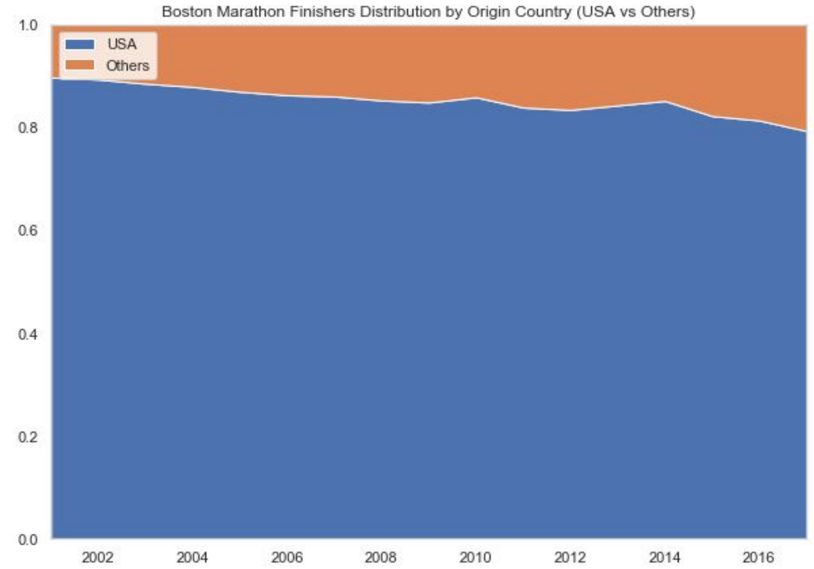
Then, we looked at the finishers distribution by age group and see if there's a major shift among them. We used the stacked area chart to visualize the percentage of each age group by year. It seems that there's a growing number of people from the age group 50s and 60s+ finished Boston Marathon year over year.

Total number of finishers by gender



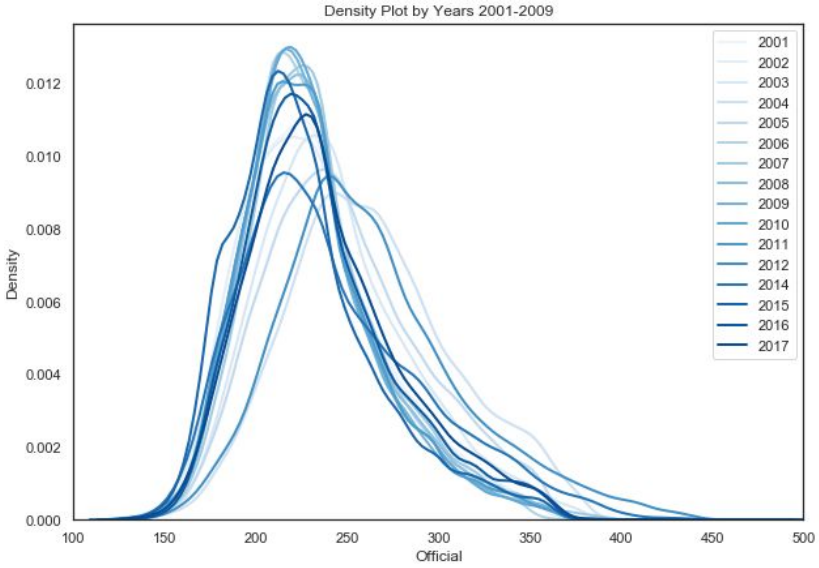
Then we looked at the ratio between female and male runners. The chart shows that the percentage of female runners were growing year over year, from 36% in 2001 to 45% in 2017.

Total number of finishers by origin country



Lastly, we looked at the ratio between runners from USA vs other countries. The chart above shows that there's a growing number of participants from international countries year over year.

What's the distribution of performance(finishing time) look like among all the runners?

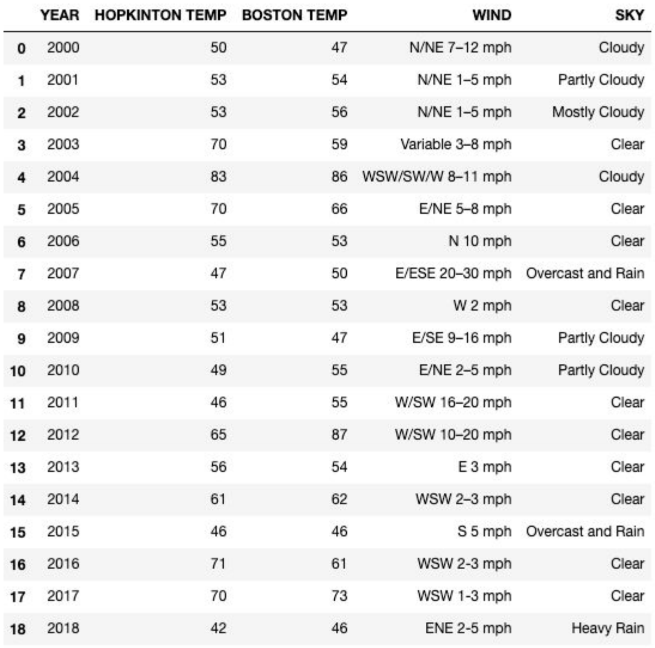


Finally, we plotted out the density plot of official finishing time by year, in the hope of seeing if there's any trend showing that runners' performance are getting better year over year. From the chart above, it's obvious that there's no trend showing that the bell-shaped density plot is skewing towards right each year. It got me thinking that, if runners are not getting better year over year, then there must be some external factors that affected their performance. This will lead to the next part of our project - correlation analysis.

Part 4: Correlation Analysis

What’s the effect of weather on runner's finishing time?

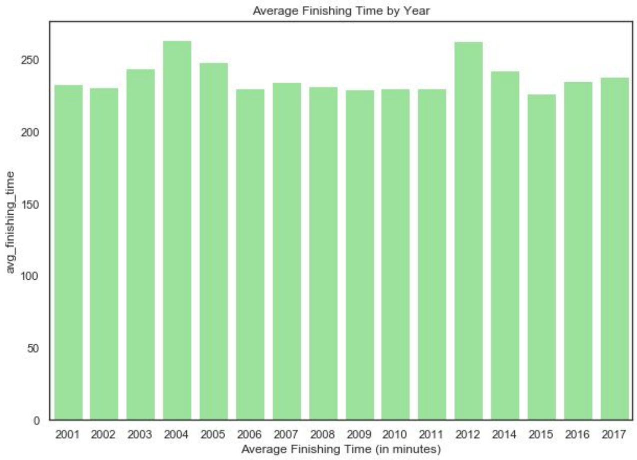
We pulled the weather data from Boston Marathon Official website. You can see that even though Boston Marathon happens around the same time every year, the weather was pretty unpredictable in terms of temperature and sky conditions.



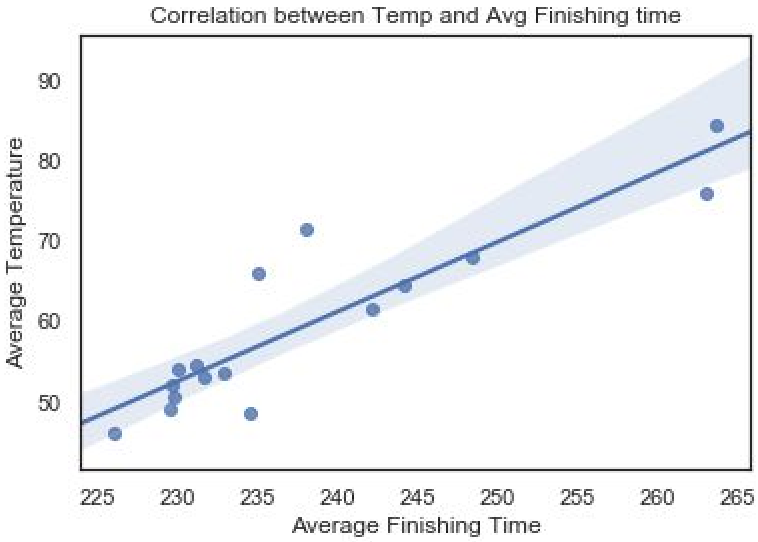
	YEAR	HOPKINTON TEMP	BOSTON TEMP	WIND	SKY
0	2000	50	47	N/NE 7–12 mph	Cloudy
1	2001	53	54	N/NE 1–5 mph	Partly Cloudy
2	2002	53	56	N/NE 1–5 mph	Mostly Cloudy
3	2003	70	59	Variable 3–8 mph	Clear
4	2004	83	86	WSW/SW/W 8–11 mph	Cloudy
5	2005	70	66	E/NE 5–8 mph	Clear
6	2006	55	53	N 10 mph	Clear
7	2007	47	50	E/ESE 20–30 mph	Overcast and Rain
8	2008	53	53	W 2 mph	Clear
9	2009	51	47	E/SE 9–16 mph	Partly Cloudy
10	2010	49	55	E/NE 2–5 mph	Partly Cloudy
11	2011	46	55	W/SW 16–20 mph	Clear
12	2012	65	87	W/SW 10–20 mph	Clear
13	2013	56	54	E 3 mph	Clear
14	2014	61	62	WSW 2–3 mph	Clear
15	2015	46	46	S 5 mph	Overcast and Rain
16	2016	71	61	WSW 2-3 mph	Clear
17	2017	70	73	WSW 1-3 mph	Clear
18	2018	42	46	ENE 2-5 mph	Heavy Rain

We used 3 attributes to look at the trend in the average finishing time - temperature, wind speed, and sky condition. For temperature, we picked max, min, and average temperature. For wind speed, we took the average wind speed. For sky condition, we grouped them into 3 levels - level 1 is clear, level 2 is cloudy, and level 3 is rainy.

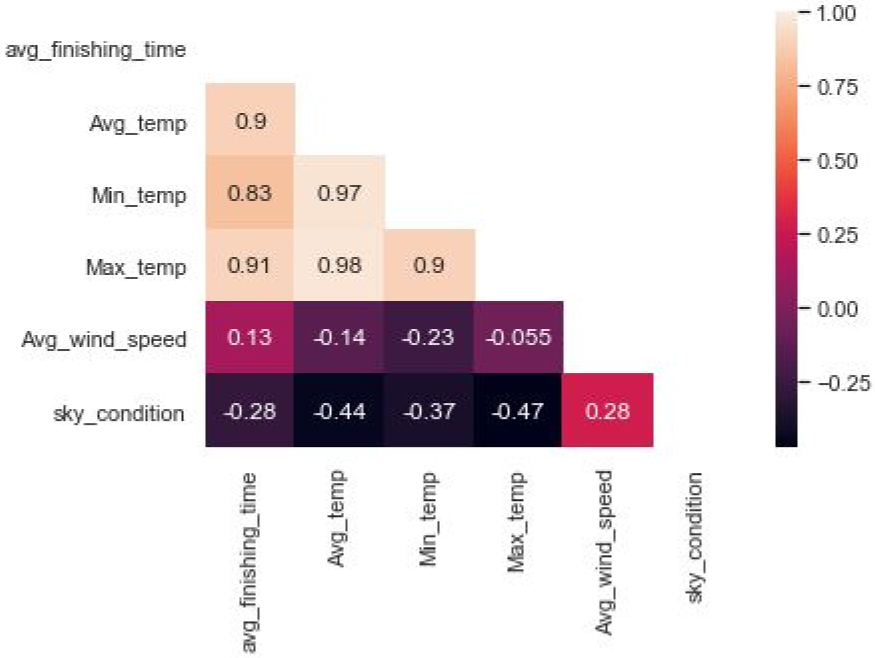
After putting together the weather data, we merged it with the full data set, and plot out the average finishing time by year -



The chart shows that there’s no clear correlation between year and the average performance. My first hypothesis is whether it has something to do with the temperature. So we plotted this correlation line chart -



This line graph indicates a clear correlation between temperature and runners performance. We further looked at the correlation with other potential factors -

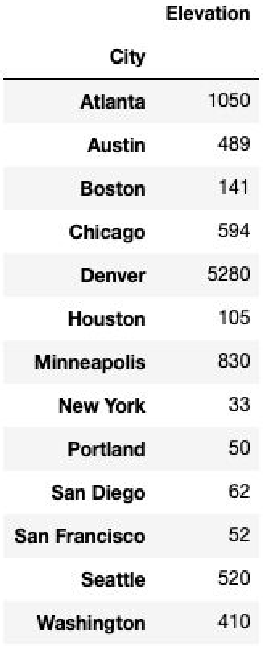


As it turned out, runners performance has a strong correlation with the temperature, but not with wind speed or sky condition. Given Boston's unpredictable weather and the wide range of high temperatures on Marathon Day, we suggest runners train for extremes so that they will be prepared for whatever they encounter, especially in hot weather.

What’s the effect of training elevation on runner's finishing time?

In this part, we attempted to understand if the elevation of where people lived and conducted their marathon training had any effect on their overall finishing times. The underlying assumption is that people would train at where they live, and that people who live and train at higher elevation would perform better than those who live and train at lower elevation.

We first filtered the databases to only keep the US runners. For the US runners, the databases contained information on which state and city each runner was from. We then aggregated the databases to determine which cities had the most runners. The year group we analyzed were 2015, 2016, and 2017. Consistently, we found the 13 cities listed in the table below had the most number of runners. Boston, being the host city, had the most number of runners of them all. The databases we used did not have city elevation information. We obtained the city elevation information through the Google search engine. We further reduced the data we analyzed to these 13 cities only.

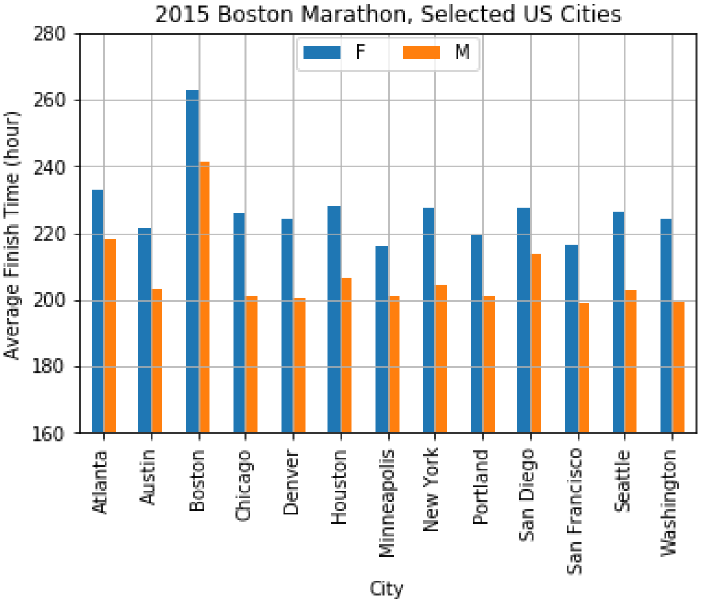


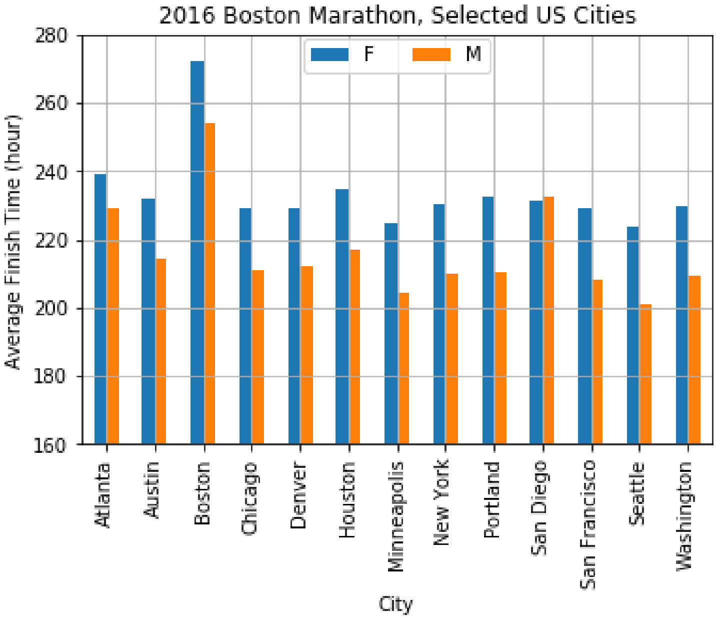
	Elevation
City	
Atlanta	1050
Austin	489
Boston	141
Chicago	594
Denver	5280
Houston	105
Minneapolis	830
New York	33
Portland	50
San Diego	62
San Francisco	52
Seattle	520
Washington	410

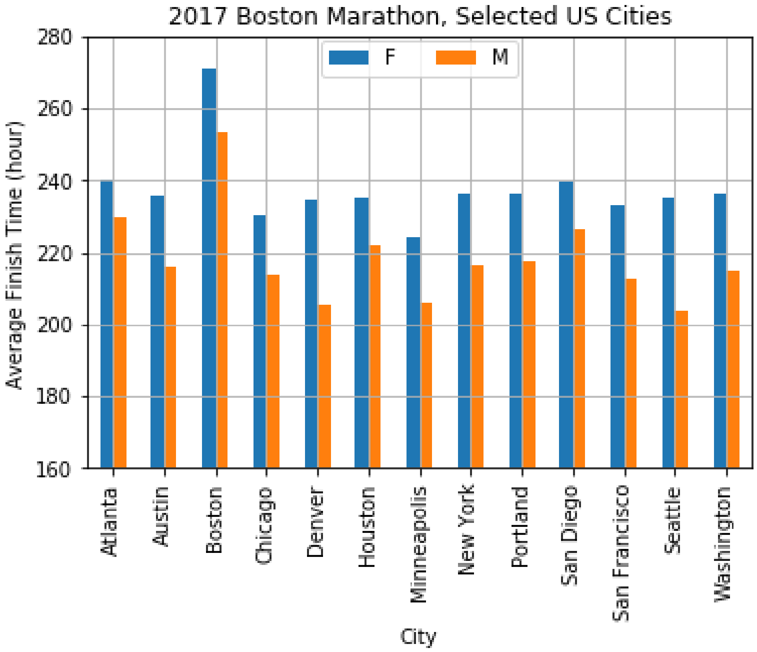
We averaged the overall finishing time by city and by gender, and we plot the results on the following bar graphs. Right off the back we noticed something rather peculiar. Boston, being the host city of the Boston Marathon, had the worst finishing time (runners taking longer to complete the race) of all cities in both the female and the male categories. We researched more into this phenomenon and theorized the cause of this skewness in the data was due to:

1. Boston Marathon is a qualified event, meaning that one has to achieve certain finishing time standard before one can participate in the race. And even after qualifying for the race, the applications have to be accepted by the event organizer as well.

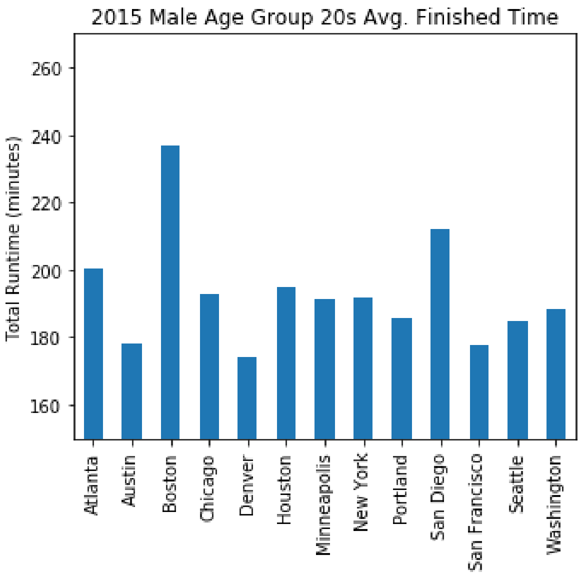
2. The restriction mentioned in 1 doesn’t apply to all Boston runners. There are several city and state sponsored programs that would allow local residents to participate in the race without qualifying for the race.
3. Non-Bostonian runners will have to make arrangements for travel and lodging. Hence, we think non-Bostonian runners are more dedicated to training for the race and getting good finishing times.

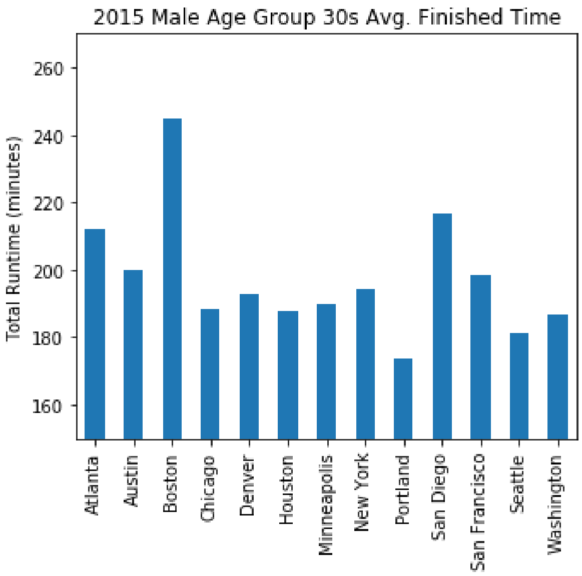

2015 Boston Marathon, Selected US Cities

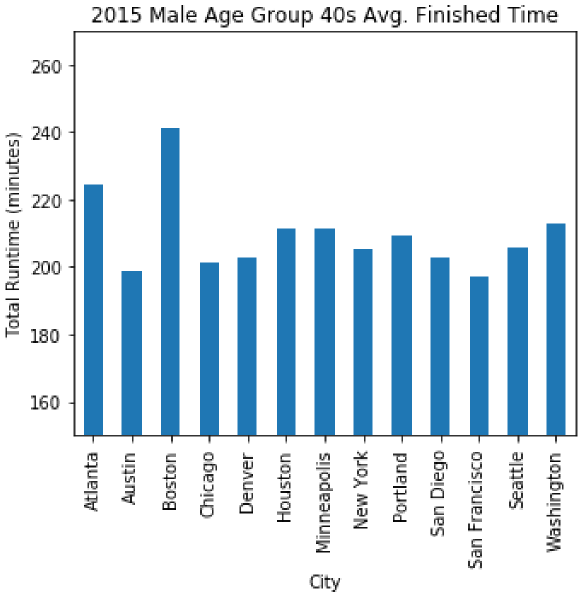

2016 Boston Marathon, Selected US Cities

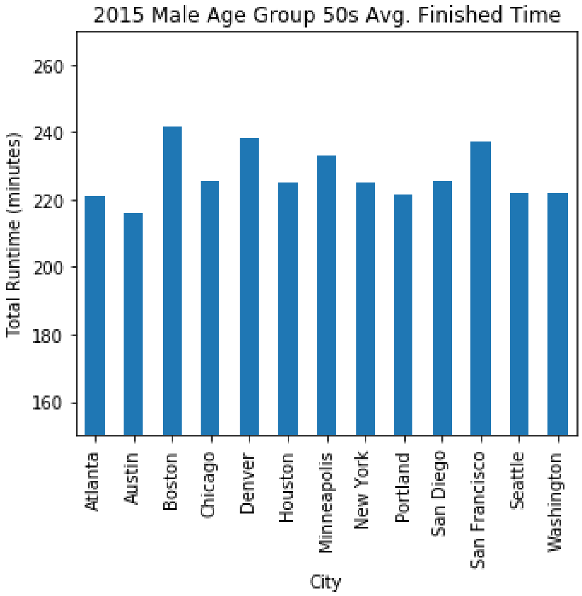

2017 Boston Marathon, Selected US Cities

We further refined the data by binning the databases into separate age groups. The databases contained a column named “Age”. We used that information and binned the runners into age groups in tens, i.e 10s, 20s, 30s, 40s, 50s, 60s, and 70s. For the refined analysis, we only kept 20s, 30s, 40s, and 50s age groups. The four bar graphs below was produced from the 2015 database. Runners analyzed were male in their 20s, 30s, 40s, and 50s. We again observed the peculiar spike from Boston. From these bar graphs, we couldn’t definitely conclude that the elevation of where people live and train has an effect on how they perform. We saw that in the Male 20s group, Denver had the best performance and Denver sits at 5280’ elevation. However, in the Male 30s group, it was actually Portland that out run all other cities. Portland sits at 50’ elevation.

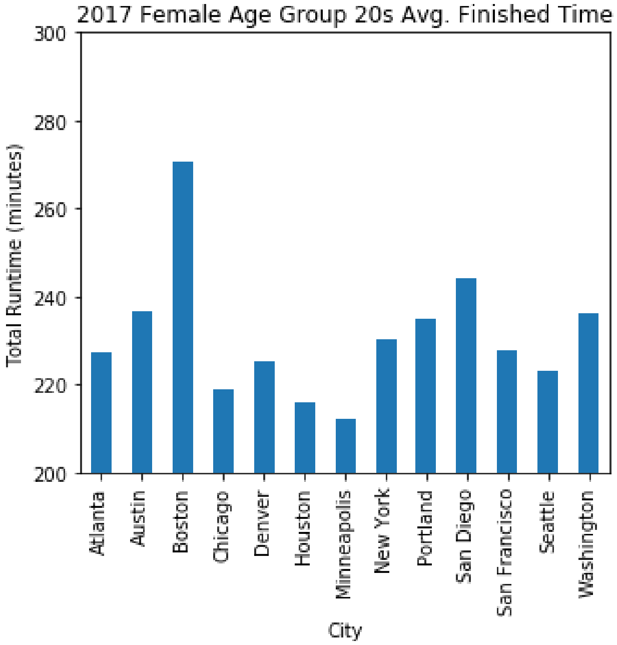


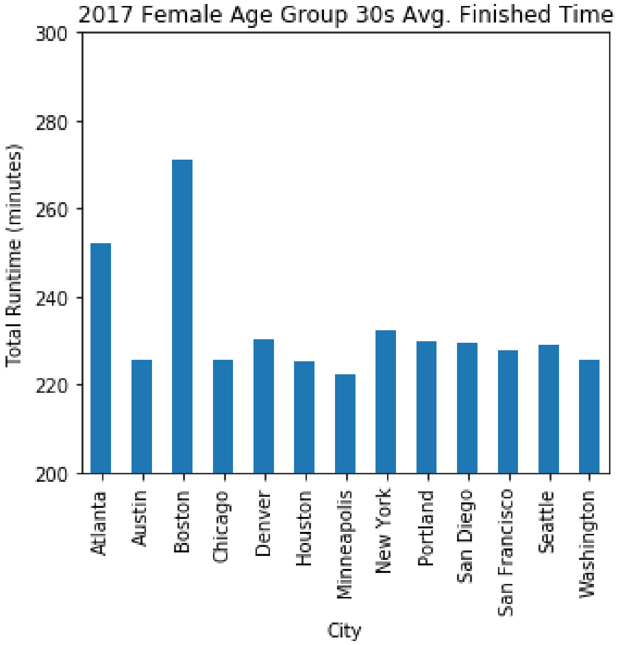


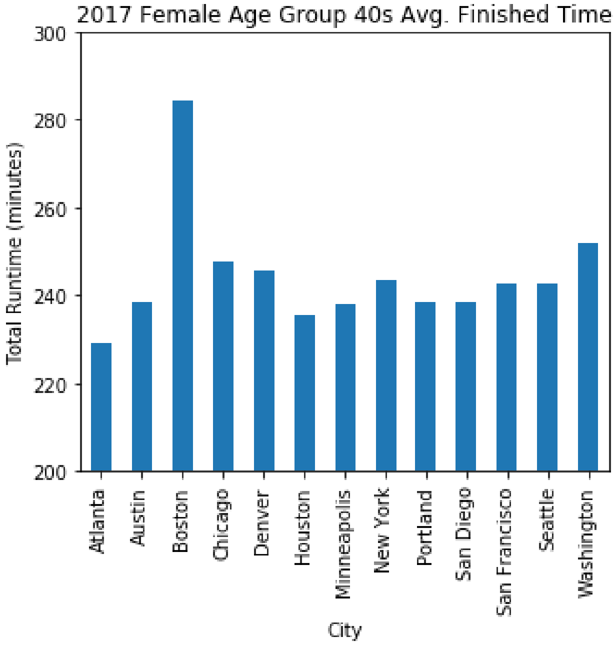


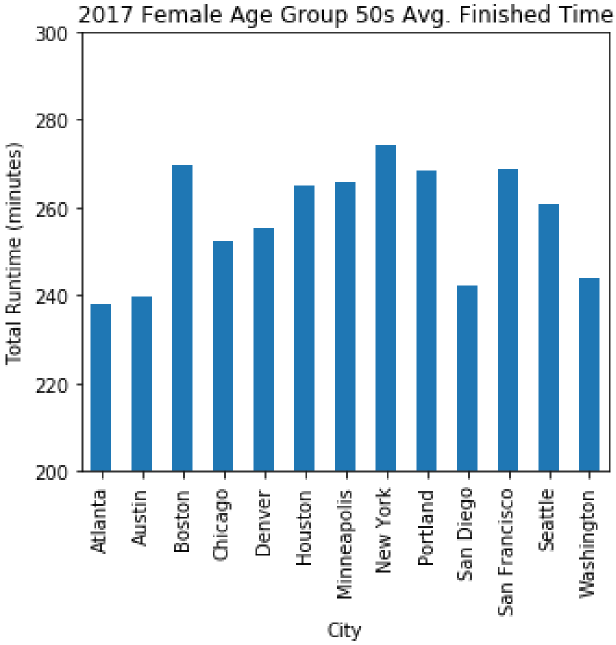


Similar inconclusiveness was observed in female runners. Following are four bar graphs from the 2017 Boston Marathon. We observed that in the Female 20s group, Minneapolis outperformed all other cities. However, in the Female 50s group, Minneapolis slacked behind San Diego by almost 30 minutes. Minneapolis is at 830' elevation, while San Diego is at 62' elevation.

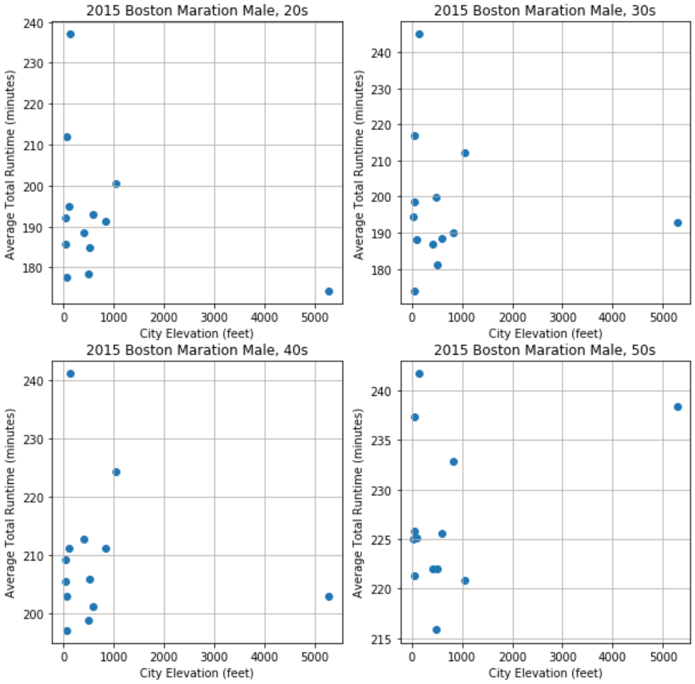




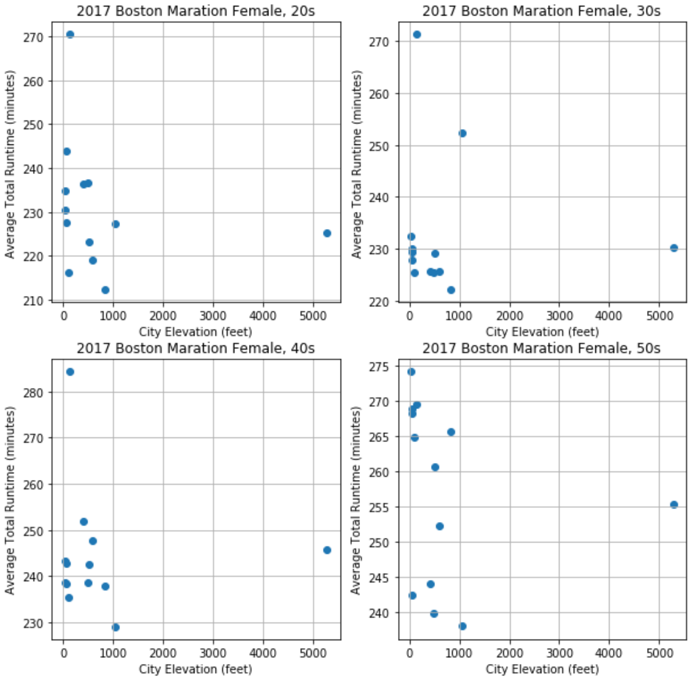




To better understand the correlation between finishing time and elevation, we scatter plot the average city finishing time against the city elevations. The four scatter plots below are from the same 2015 Male groups.



The four scatter plots below are from the same 2017 Female groups.

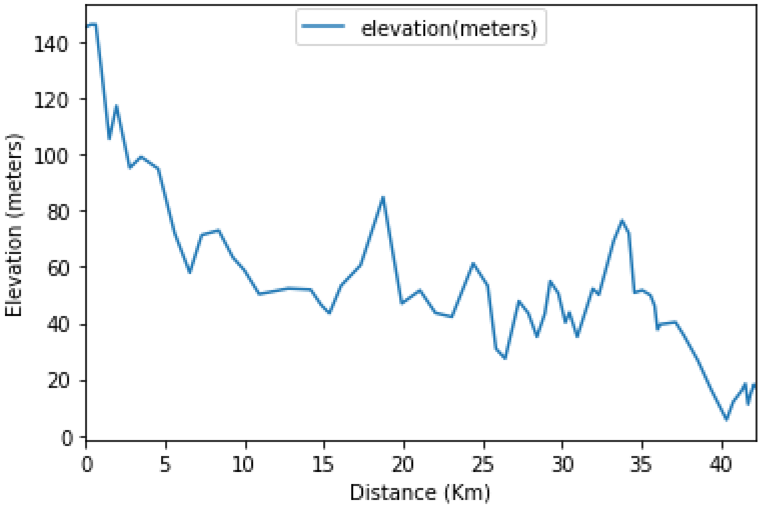


We observed that at lower elevation, there was quite a spread in the finishing time, and provided no conclusive evidence that people who live and train at higher elevation do indeed perform better than those who live and train at lower elevation.

What’s the effect of course elevation on runner's finishing time?

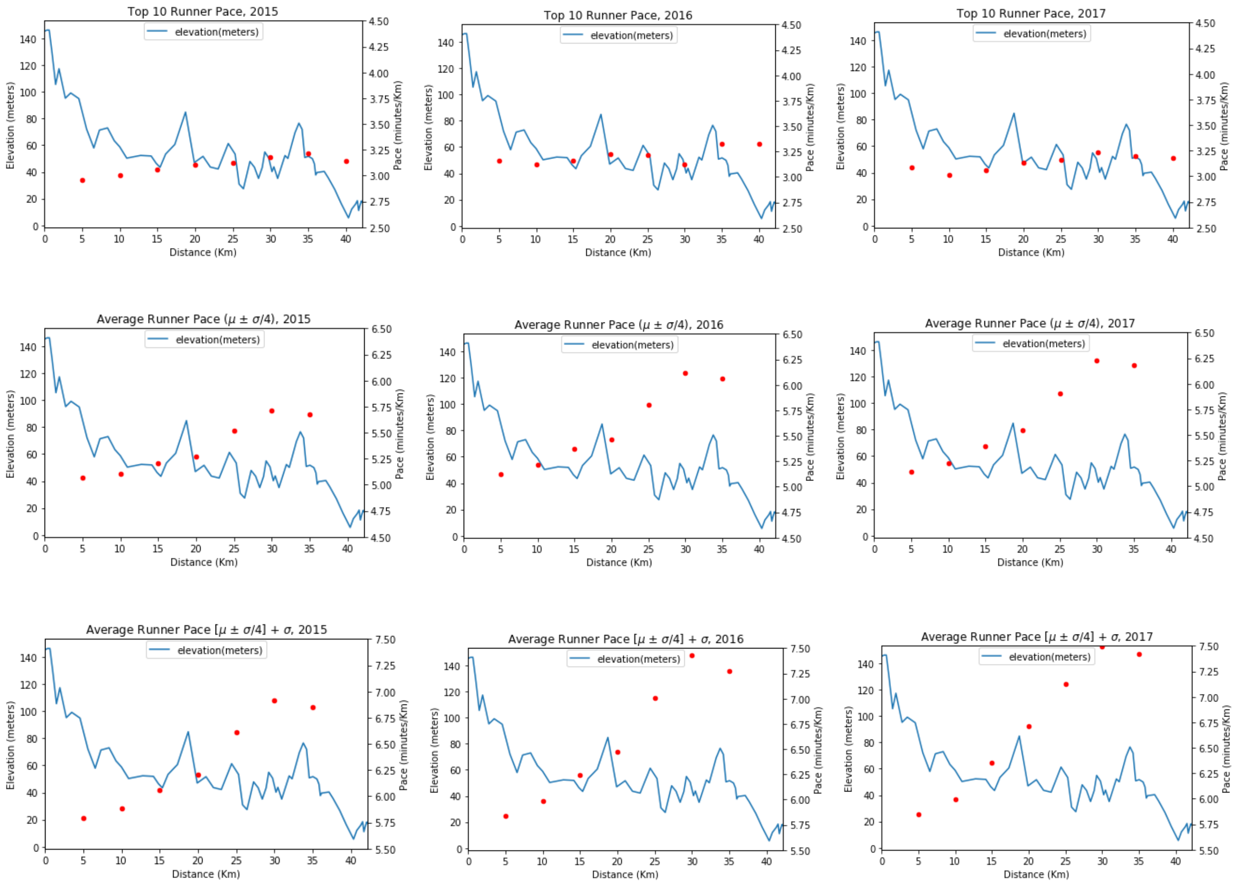
In this part, we looked at how runners responded to the course elevation change. We were particularly interested in comparing the pace strategy of regular runners with elite runners.

We obtained the course elevation data from Kaggle.com. The data contains four columns. The first column is the distance from the start. The second column is the elevation at that point. The third column and the fourth column are cumulative ascend and descend, respectively. The data had several repeated rows of the same information. We used the unique() method to filter out repeated rows and plot the course elevation in the graph below.



Next, we manipulated 2015, 2016, and 2017 databases to calculate runner’s pace throughout the course. In each database, time hacks were provided for 5K, 10K, 15K, …, and 40K distance mark. We calculated the pace by taking the difference between each time hack and divided by 5K. For example, the pace at 5K would be the time hack at 5K divided by 5. The pace at 30K would be the time hack at 30K minus the time hack at 25K and divided by 5.

For each year, we selected three groups of runners to compare. The first group is the “elite’ group. They were the top 10 finishers from each year. The second group is the “average” group. They were the runners, whose finishing time fell within 0.25 standard deviation from the mean. The last group is the “slow” group. They were the runners, whose finishing time fell within 1.0 standard deviation above the mean ± 0.25 standard deviation. We plotted the average pace of each group against the course elevation. These plots are shown as followed.


Top 10 Runner Pace, 2015

Top 10 Runner Pace, 2016

Top 10 Runner Pace, 2017

Average Runner Pace ($\mu \pm \sigma/4$), 2015

Average Runner Pace ($\mu \pm \sigma/4$), 2016

Average Runner Pace ($\mu \pm \sigma/4$), 2017

Average Runner Pace [$\mu \pm \sigma/4$] + σ, 2015

Average Runner Pace [$\mu \pm \sigma/4$] + σ, 2016

Average Runner Pace [$\mu \pm \sigma/4$] + σ, 2017

For the nine plots above, the first row are "elite" runners. The second row are "average" runners. The last row are "slow" runners. The first column is from 2015. The second column is from 2016. The last column is from 2017. In each plot, the blue line represents the course elevation. Its scale is on the left y-axis. The red dots represent the average pace of the group of runners. Its scale is on the right y-axis. Key thing to note here is that the right y-axis spans 2 minutes per kilometer.

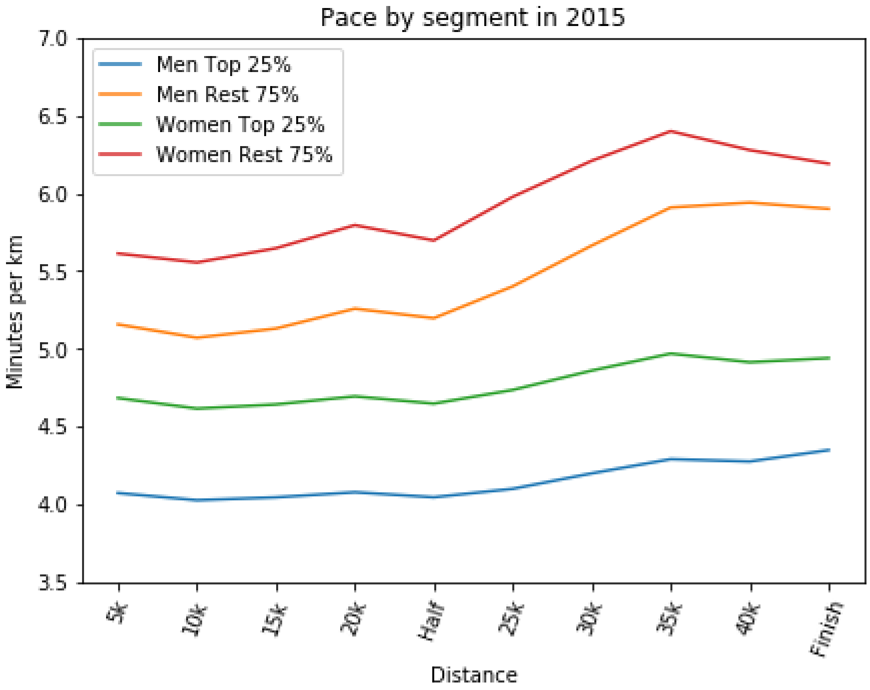
From these plots, we noted that "elite" runners tend to maintain their paces throughout the course regardless of the course elevation change. There are some small variations in their paces. But on average the pace stayed fairly constant at around 3.25 minutes per kilometer throughout the course. On the contrary, "average" and "slow" runners tend to run faster in the beginning and gradually slowed down throughout the course. This observation is suggestive that "average" and "slow" runners attempted to capitalize on initial descent of the course elevation. However, this strategy was not adopted by "elite" runners and could result in slow finishing time. Hence, we recommend to run like "elite" runners and keep a constant pace throughout the race and not react to the course elevation change.

Part 5: Cluster Analysis

In this part, we used the segment pacing information. Before 2013, the data only have finish time and we cannot calculate the pacing information. Therefore, we used the data from 2014 to 2017. Also, the data of 2014 do not have the time at 15k. This is why we started the pacing analysis below from 2015.

Quantile Analysis

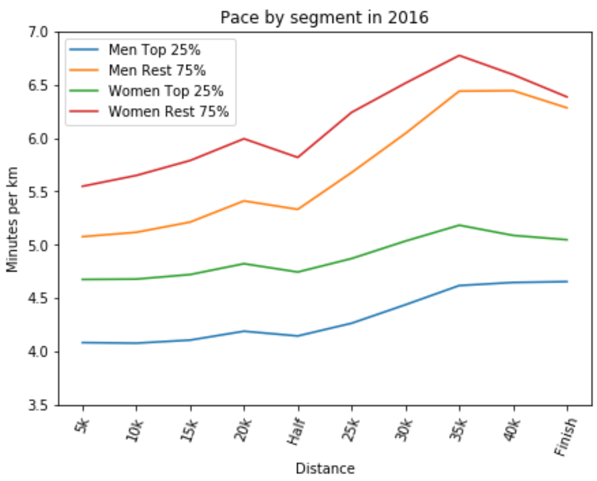
First, we compared the pace by distance between the top 25% runners and the rest of the runners in 2015. Although both groups show the drop of the pace in the second half, the pace of top 25% runners are more stable than the rest of the runners.

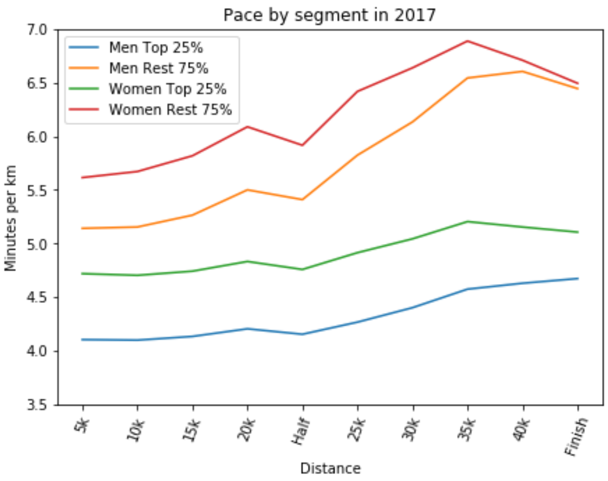


Second, we did the same analysis using the data of 2016 and 2017 to see if the same trend can be observed. As shown in the following page, we can see the same trend as in 2015. However, the drop is more apparent in 2016 and 2017 for both groups.

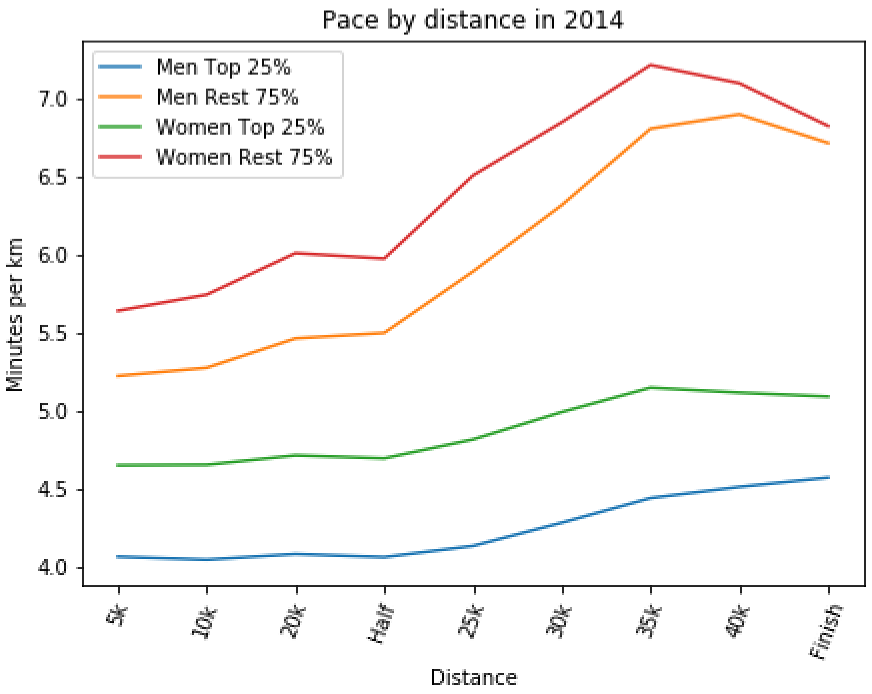
According to the weather information from the website of Boston Athletic Association shown below, it was hotter in 2016 and 2017. At the time the race started, the temperature was 61F and the it was 71F when the first runner finished in 2016. On the contrary, it was 46F at the time of the race started, and it was also 46F when the first runner finished in 2015, which is colder than the average temperature of Boston in April.

As we mentioned in part 4, the temperature has a huge impact on the finishing time. It also has a huge impact on segment pacing.


Pace by segment in 2016


Pace by segment in 2017

The data in 2014 do not have time at 15k. Therefore, the figure is a little different from the previous ones. However, we can observe the same trend as in 2016 and 2017. In 2014, the temperature was higher than 2015.


Pace by distance in 2014

WEATHER CONDITIONS

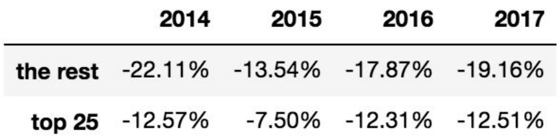
YEAR	HOPKINTON TEMP*	BOSTON TEMP**	WIND	SKY
2014	61	62	WSW 2–3 mph	Clear
2015	46	46	Calm	Overcast and Rain
2016	71	61	WSW 2-3 mph	Clear
2017	70	73	WSW1-3 mph	Clear

*Based on start of Wave One

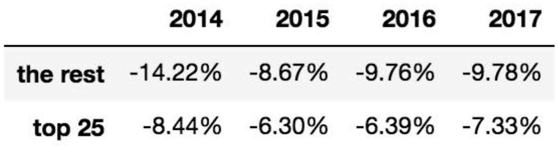
**Based on winner of men's race

Source: https://www.baa.org/races/boston-marathon/results/history

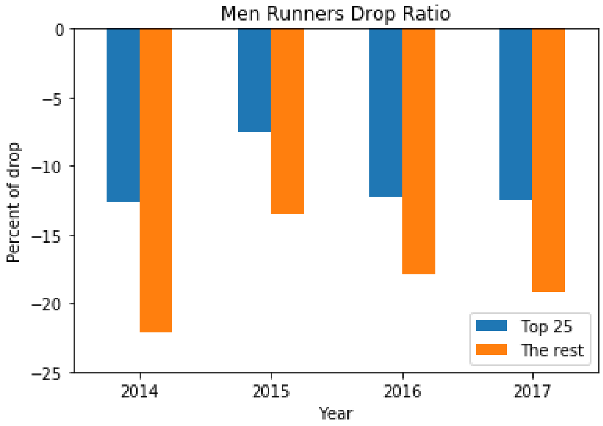
We also calculated the “drop ratio”, which shows how the pace dropped by comparing the pace at the last segment with the pace at the half. The left figure is the drop ratio from 2015 to 2017 for men runners and the right is the one for women runners.

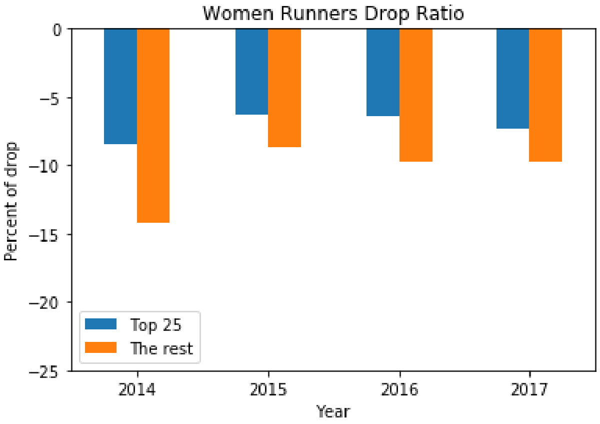


	2014	2015	2016	2017
the rest	-22.11%	-13.54%	-17.87%	-19.16%
top 25	-12.57%	-7.50%	-12.31%	-12.51%



	2014	2015	2016	2017
the rest	-14.22%	-8.67%	-9.76%	-9.78%
top 25	-8.44%	-6.30%	-6.39%	-7.33%





Overall, the drop ratio of top 25% is smaller than the rest of the group. Also, we observe that men runners are more likely to drop the pace in the second half.

From these four-year data above, we can say the stable pacing is a good strategy for Boston Marathon. Especially, when the temperature is high, it is more likely to drop the pace in the second half. Runners should save the pace to avoid the big drop in the second half, especially when the temperature is high.

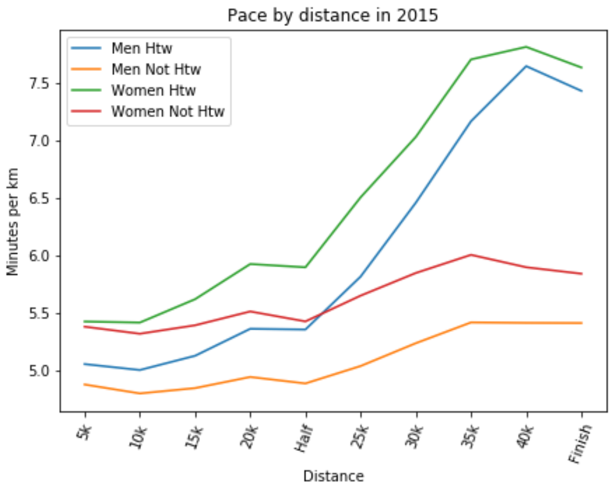
Hit the wall analysis

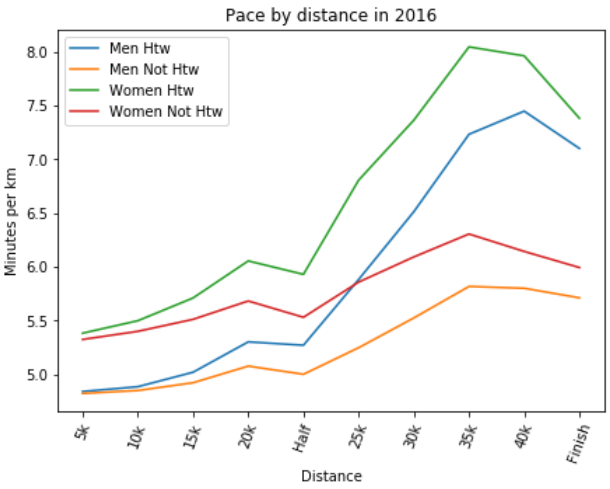
Next, we focus more on the drop in the second half to understand the “wall” so that we find out a strategy how to practice before the race. Also, it would be beneficial for runners to understand the wall to avoid the overpace.

In order to do this analysis, we defined “Hit the wall” as follows.

Our definition:
if any of the 5k pace in the second half > first half pace * 1.33, we consider the person hit the wall.

With this definition, we divided the runners into two groups: Htw and Not Htw. We just show two figures since the trend of is almost the same.

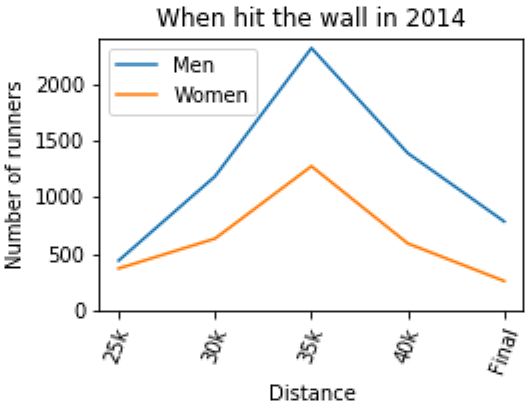


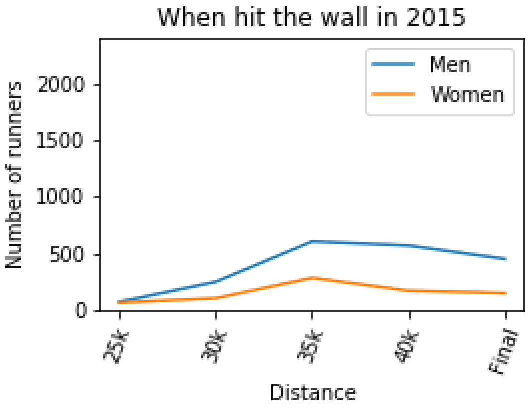


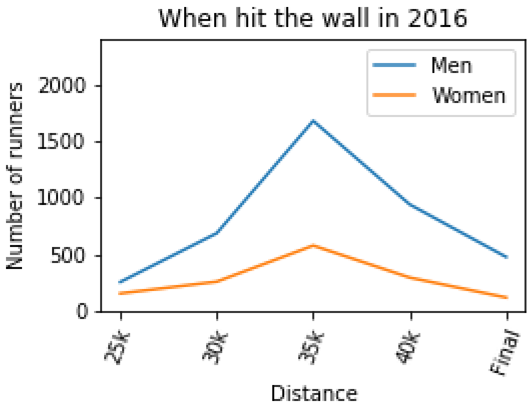
The following figures shows in which segment runners hit the wall for the first time in the second half in 2014, 2015, 2016, and 2017. The biggest wall was the segment between 30k and 35k as it is often said. There are two uphills between 30k and 35k, which makes much tougher for runners.

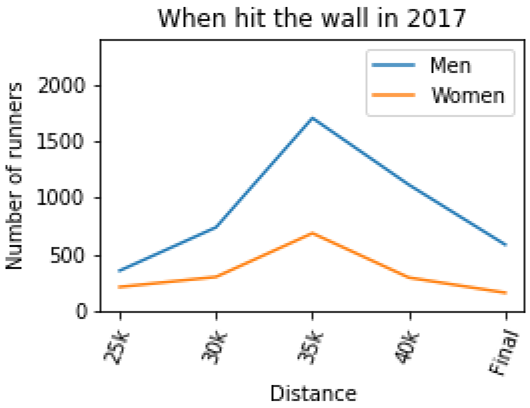
As observed in the previous analysis, more runners hit the wall in 2014, 2016 and 2017 than 2015. We can see the effect of high temperature in this analysis as well.

From the hit the wall analysis, in addition to the pacing, it is important to run 30k or 20 miles periodically before the race. Runners cannot overcome the wall of 30k without running 30k beforehand.









Conclusion

Here are our research questions we mentioned before.

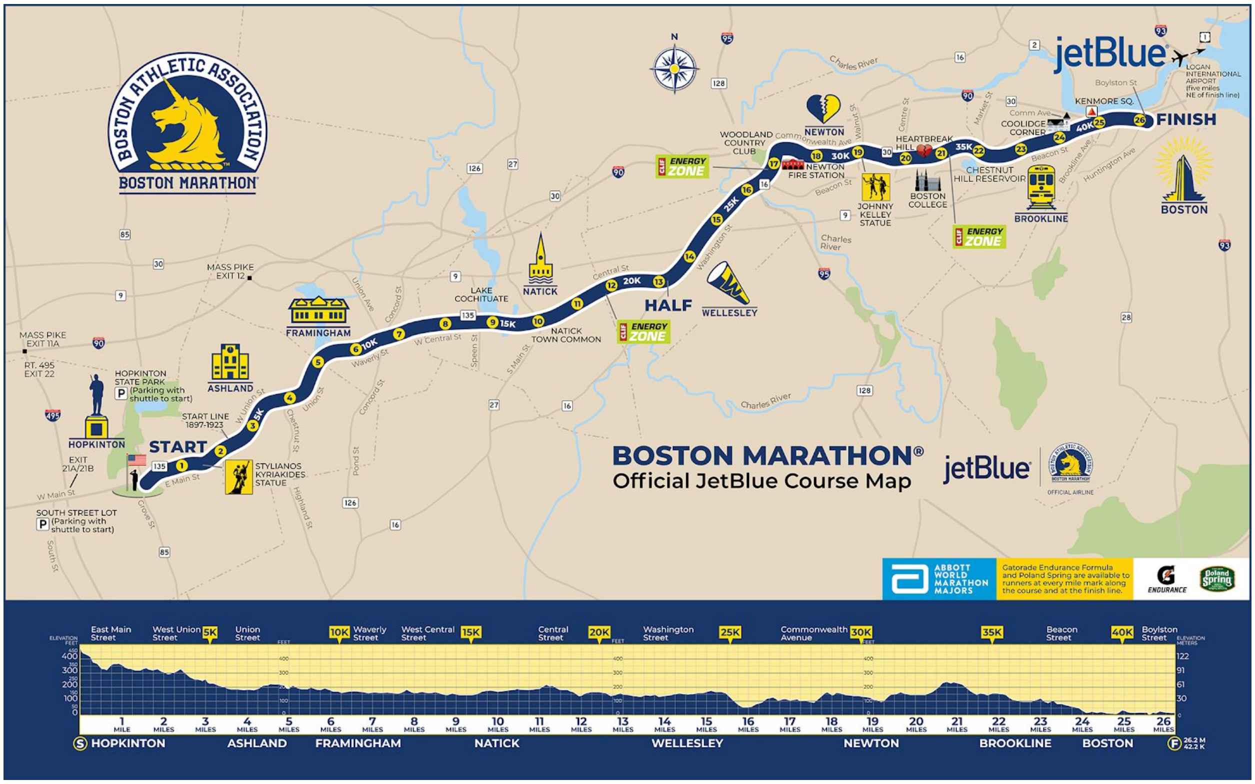
- How does weather condition affect the overall finishing time for Boston Marathon?
- How does the elevation change affect runners' pace?
- What’s the good pacing strategy for Boston Marathon?

From our analysis, our recommendations for the Boston Marathon runners are as follows..

1. Prepare the big wall around 30k (20 miles). In order to prepare for this, runners should practice running 30k (20 miles) before the race. Ideally, it is better to practice the course where there are several uphills around 30k (20 miles).
2. Try to keep the same pace. Overpace in the first half causes the huge drop in the second half. Especially, men tend to be positive split. Therefore, we recommend to save the pace in the first half so that runners do not drop the pace in the second half.
3. Boston Marathon has uphills and downhills. However, we found that keeping the stable pace is a good strategy for runners.
4. When thinking of the right pace, runners should consider temperature. As we mentioned, temperature has strong correlation with the finishing time. Runners usually decide pacing plan before the race. If the weather report forecasts high temperature on race day, it is better to set the pacing plan more conservatively.

Appendix

Boston Marathon Official Course Map


BOSTON ATHLETIC ASSOCIATION
BOSTON MARATHON
BOSTON MARATHON®
Official JetBlue Course Map
jetBlue
OFFICIAL AIRLINE
START
HALF
FINISH
HOPKINTON
ASHLAND
FRAMINGHAM
NATICK
WELLESLEY
NEWTON
BROOKLINE
BOSTON
HOPKINTON STATE PARK
P (Parking with shuttle to start)
START LINE 1897-1923
STYLIANOS KYRIAKIDES STATUE
SOUTH STREET LOT
P (Parking with shuttle to start)
MASS PIKE EXIT 11A
RT. 495 EXIT 22
EXIT 21A/21B
MASS PIKE EXIT 12
LAKE COCHITUATE
NATICK TOWN COMMON
WOODLAND COUNTRY CLUB
NEWTON FIRE STATION
JOHNNY KELLEY STATUE
HEARTBREAK HILL
BOSTON COLLEGE
CHESTNUT HILL RESERVOIR
COOLIDGE CORNER
KENMORE SQ.
LOGAN INTERNATIONAL AIRPORT (five miles NE of finish line)
ENERGY ZONE
5K
10K
15K
20K
25K
30K
35K
40K
Charles River
Boylston St
Comm Ave
Commonwealth Ave
Beacon St
Brookline Ave
Huntington Ave
Centre St
Walnut St
Market St
Washington St
Central St
W Central St
Waverly St
Union Ave
Concord St
Speen St
S Main St
Union St
W Union St
Chestnut St
Pond St
Highland St
E Main St
W Main St
Grove St
South St
N
ABBOTT WORLD MARATHON MAJORS
Gatorade Endurance Formula and Poland Spring are available to runners at every mile mark along the course and at the finish line.
ENDURANCE
Poland Spring
ELEVATION FEET
ELEVATION METERS
East Main Street
West Union Street
Union Street
Waverly Street
West Central Street
Central Street
Washington Street
Commonwealth Avenue
Beacon Street
Boylston Street
400
300
200
100
0
122
91
61
30
0
1 MILE
2 MILES
3 MILES
4 MILES
5 MILES
6 MILES
7 MILES
8 MILES
9 MILES
10 MILES
11 MILES
12 MILES
13 MILES
14 MILES
15 MILES
16 MILES
17 MILES
18 MILES
19 MILES
20 MILES
21 MILES
22 MILES
23 MILES
24 MILES
25 MILES
26 MILES
S HOPKINTON
ASHLAND
FRAMINGHAM
NATICK
WELLESLEY
NEWTON
BROOKLINE
BOSTON
F 26.2 M 42.2 K