



Boston Marathon Run Times: 2002 - 2014

Ash Mohan, Nathaniel Medina, Nicolas Pappas, Rick Yang
Nov 28th, 2017

Personal Connection

Iconic Annual Event

Qualifying Standards

Worldwide Participation



Related Works

- Runners of the Boston Marathon 2017
- A Data Analysis of the 2016 Boston Marathon Finishers
- Finishers Boston Marathon 2015, 2016, 2017



- Retrieved from Github
- 14 distinct csv files, separated by year
- Years covered: 2001 - 2014



Average run times are increasing due to ...

- Aggregating the data
- Selecting relevant columns
- Creating additional columns
- Changing columns to the correct data type
- Removing missing (N/A) values



Data Cleaning Overview

Before

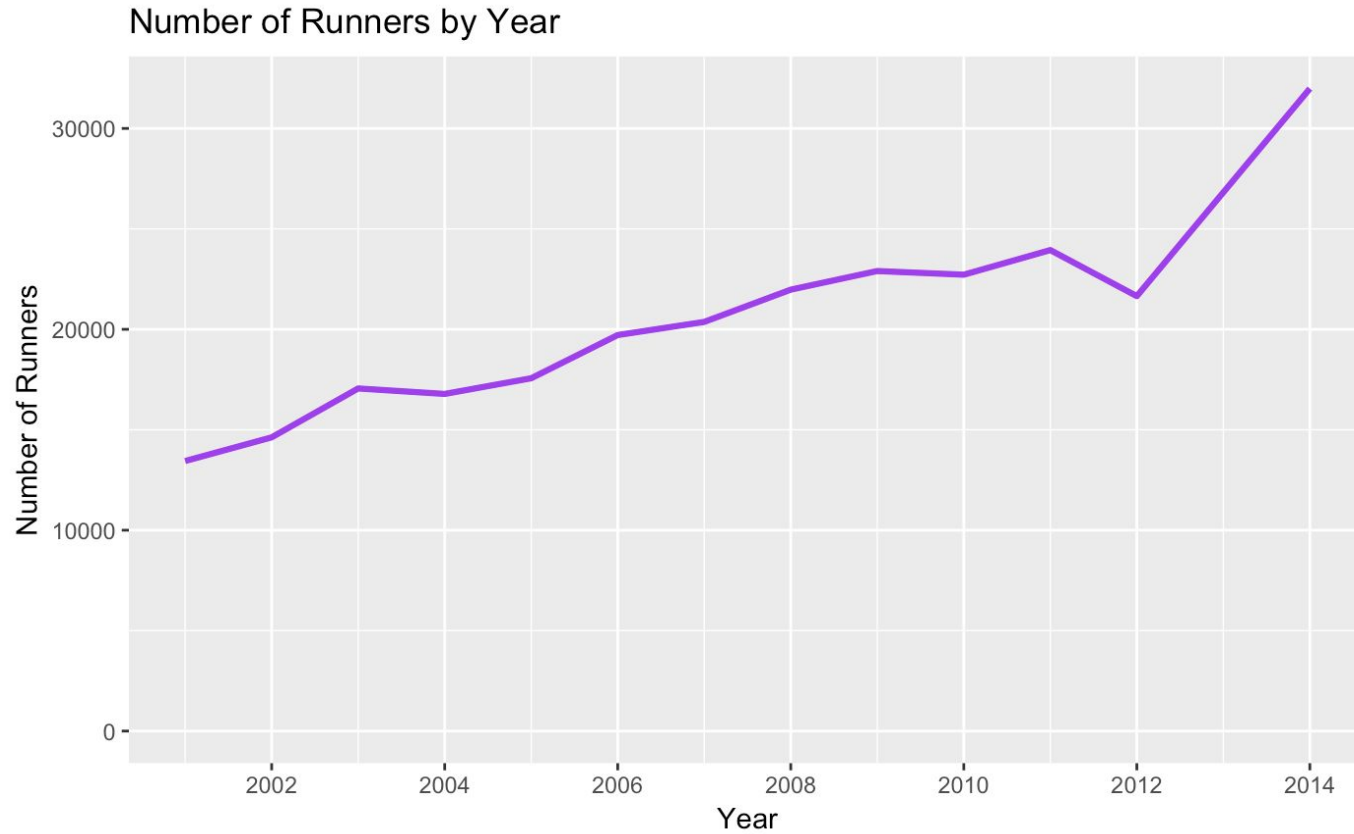
	X10k	name	division	X25k	gender	age	official	bib	genderdiv	ctz	X35k	overall	pace	state	X30k	X5k	half	X20k	country	city	X40k
1	17.37	Yamamoto, Hiroyuki	8	47.67	M	47	85.25	W1	8		71.40	8	3.27		59.18	8.02	39.72	37.65	JPN	Fukuoka	80.43
2	32.58	Jeptoo, Rita	1	82.43	F	33	138.95	F1	1		116.37	21	5.30		99.33	16.22	69.47	65.83	KEN	Eldoret	132.10
3	16.62	Van Dyk, Ernst F.	1	45.80	M	41	80.60	W2	1		67.42	1	3.08		56.45	7.75	38.03	36.10	RSA	Paarl	76.10
4	32.57	Dibaba, Mare	3	82.43	F	24	140.58	F2	3		116.37	27	5.37		99.33	16.20	69.47	65.83	ETH	Shoa	132.95
5	17.12	Hokinoue, Kota	2	46.37	M	40	81.23	W3	2		67.83	2	3.10		57.03	8.02	38.60	36.58	JPN	Nogata Fukuoka	76.72
6	32.58	Sumgong, Jemima Jelagat	4	82.45	F	29	140.68	F3	4		116.37	28	5.37		99.33	16.22	69.47	65.83	KEN	Nandi	132.95
7	17.65	Hug, Marcel E.	4	47.67	M	28	84.65	W4	4		70.23	4	3.23		58.60	8.38	39.72	37.65	SUI	Neuenkirch	79.83
8	30.48	Geneti, Markos	5	76.95	M	29	129.83	5	5		107.47	5	4.97		92.52	15.17	64.85	61.62	ETH	Addis Ababa	122.80
9	17.12	Soejima, Masazumi	3	46.37	M	43	81.23	W6	3		67.83	3	3.10		57.03	8.00	38.60	36.60	JPN	Isahaya	76.70
10	30.48	Hall, Ryan	20	77.68	M	31	137.83	6	20		112.27	20	5.27	CA	94.78	15.15	65.23	61.78	USA	Redding	129.83
11	32.58	Cherop, Sharon	8	82.45	F	30	143.00	F6	8		116.92	40	5.47		99.35	16.22	69.47	65.85	KEN	Marakwet	134.65
12	30.52	Chebet, Wilson	2	76.97	M	28	128.80	7	2		107.47	2	4.92		92.52	15.18	64.87	61.63	KEN	Marakwet	121.95
13	18.13	Lemeunier, Denis Sr.	11	49.17	M	49	86.95	W7	11		72.60	11	3.32		60.85	8.40	41.18	39.03	FRA	Taule	82.17
14	32.58	Kilel, Caroline	17	85.55	F	33	152.07	F7	17		124.22	121	5.80		104.62	16.22	71.10	67.15	KEN	Bomet	143.50
15	17.12	Cassidy, Josh R.	13	48.88	M	29	87.40	W8	13		72.83	13	3.35	ON	60.82	8.02	40.20	38.02	CAN	Port Elgin	82.38
16	33.93	Linden, Desiree	10	85.03	F	30	143.90	F8	10		119.68	46	5.50	MI	102.40	16.92	71.68	67.92	USA	Rochester Hills	136.47

Data Cleaning Overview

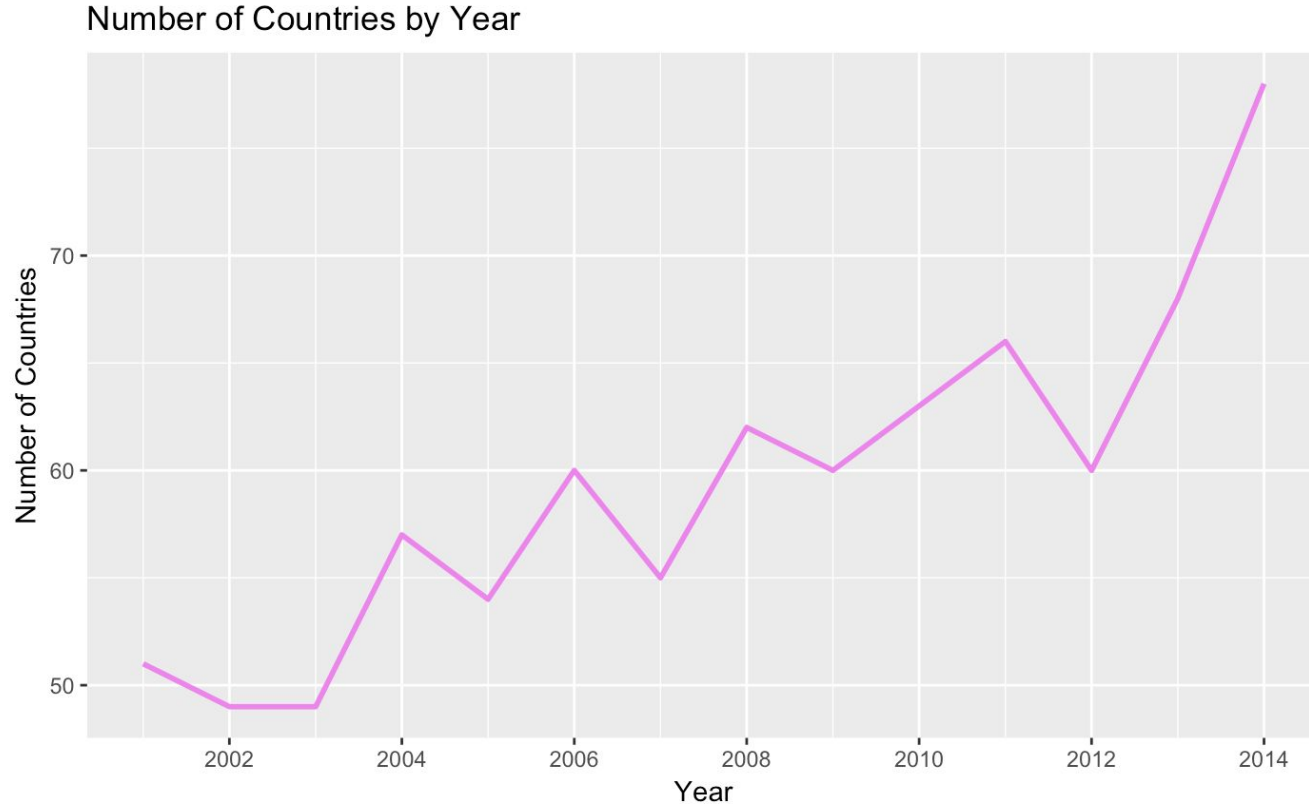
After

	name	city	gender	age	official	country	year
1	Cassidy, Josh R.	Toronto	M	28	90.90	CAN	2013
2	Korir, Wesley	Kenya	M	30	132.50	KEN	2013
3	Desisa, Lelisa	Ambo	M	23	130.37	ETH	2013
4	Fearnley, Kurt H.	Hamilton	M	32	88.43	AUS	2013
5	Hokinoue, Kota	Iizuka	M	39	87.22	JPN	2013
6	Gebremariam, Gebregziabher	Tigray	M	28	130.47	ETH	2013
7	Soejima, Masazumi	Fukuoka	M	42	90.02	JPN	2013
8	Geneti, Markos	Addis Ababa	M	28	132.73	ETH	2013
9	Schabert, Krige	Cedartown	M	49	91.78	USA	2013
10	Merga, Deriba	Addis Ababa	M	32	141.67	ETH	2013
11	Matebo, Levy	Kitale	M	23	135.70	KEN	2013
12	Van Dyk, Ernst F.	Paarl	M	40	87.20	RSA	2013
13	Hamerlak, Tomasz	Bystra	M	37	91.78	POL	2013
14	Hug, Marcel E.	Nottwil	M	27	88.32	SUI	2013
15	Chumba, Dickson	Nandi	M	26	134.13	KEN	2013
16	Botello Jimenez, Rafael	Manlleu	M	34	95.47	ESP	2013

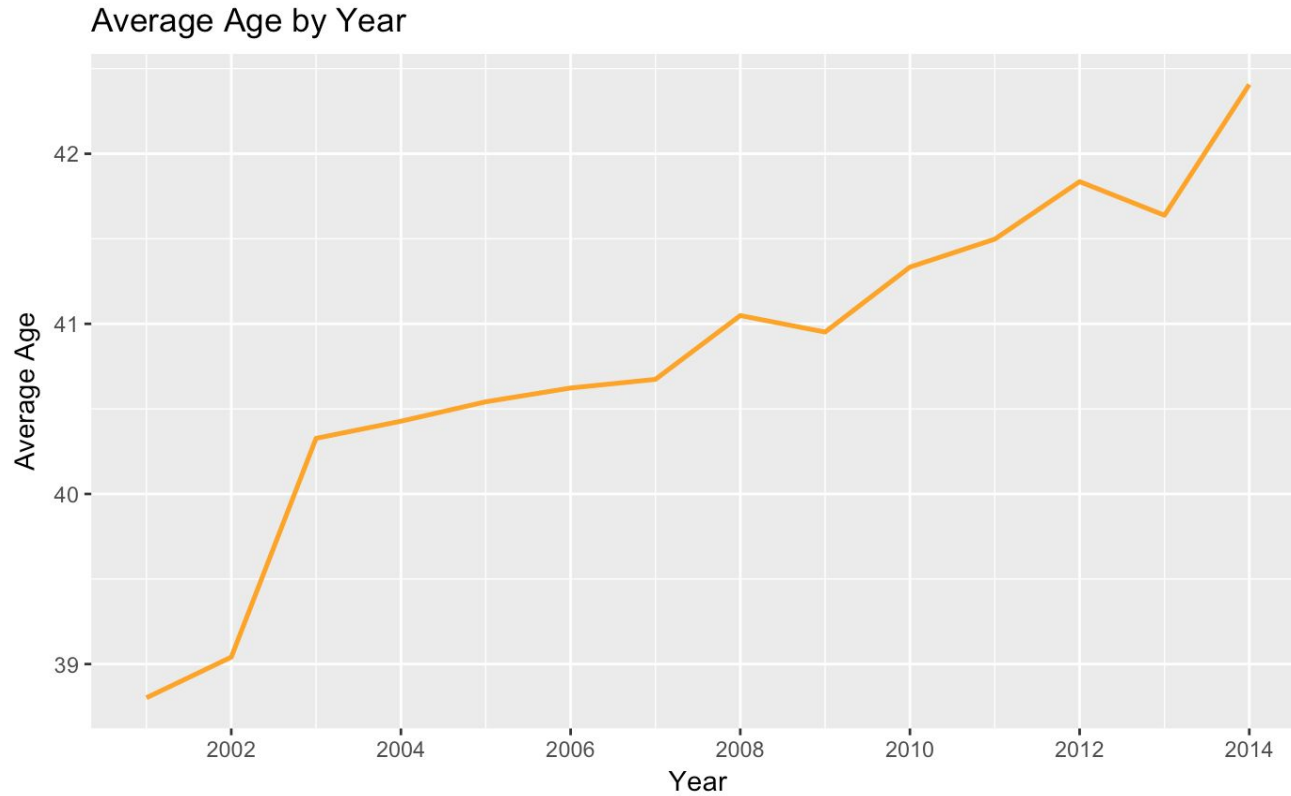
Visualizations (Participants)



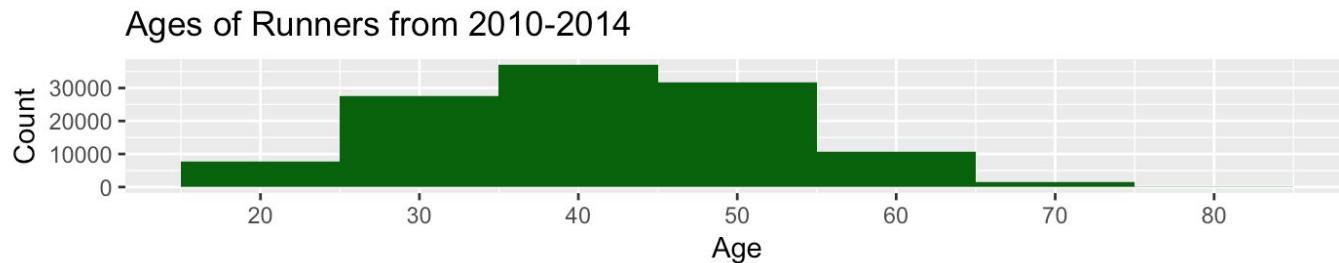
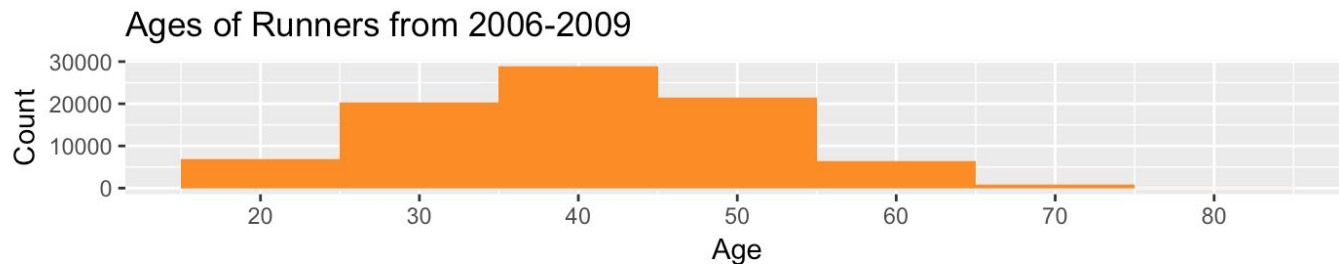
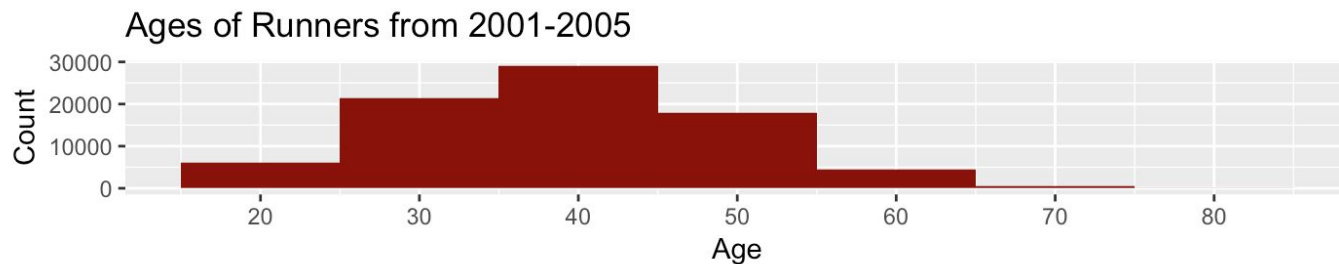
Visualizations (Countries)



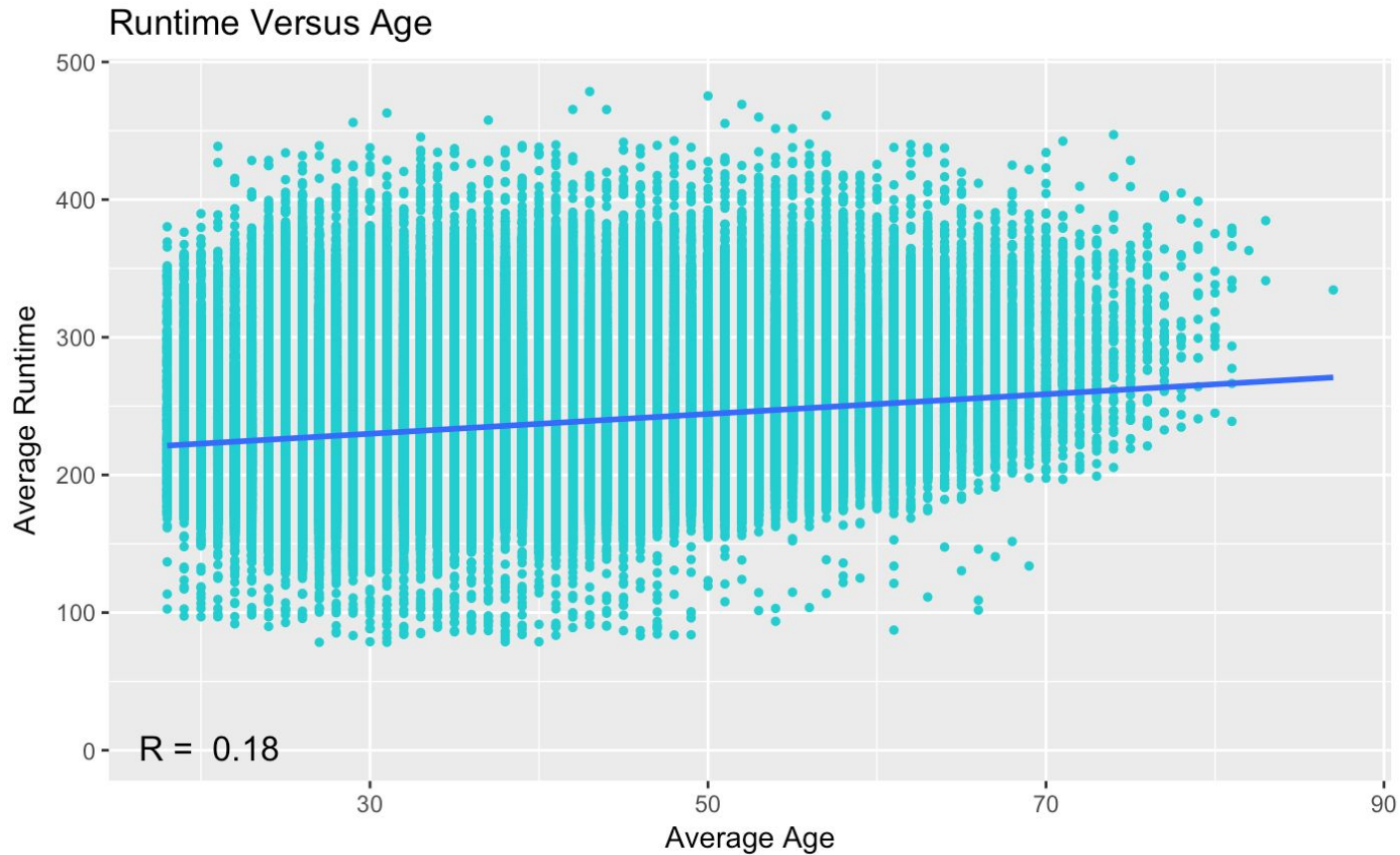
Visualizations (Age)



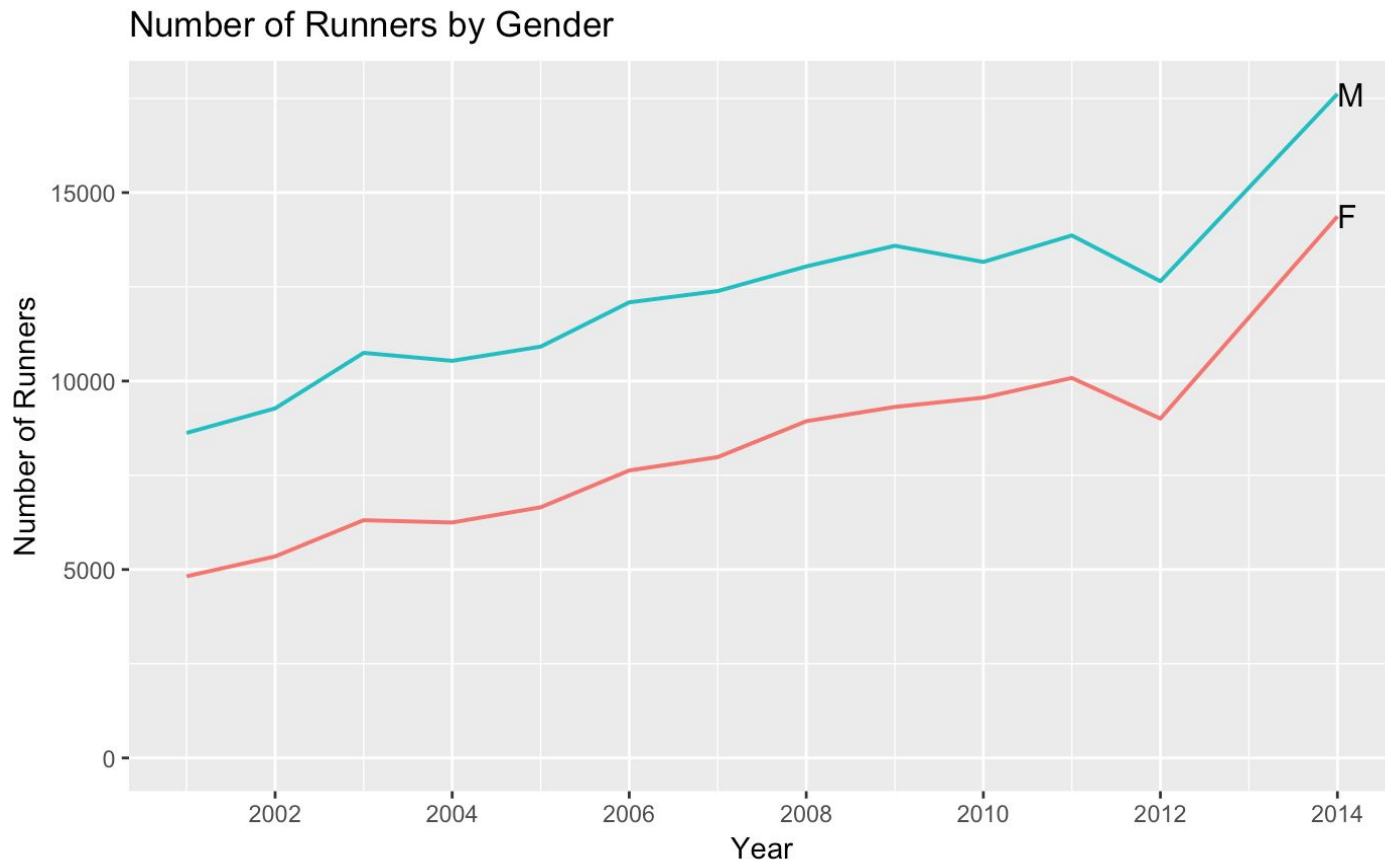
Visualizations (Age)



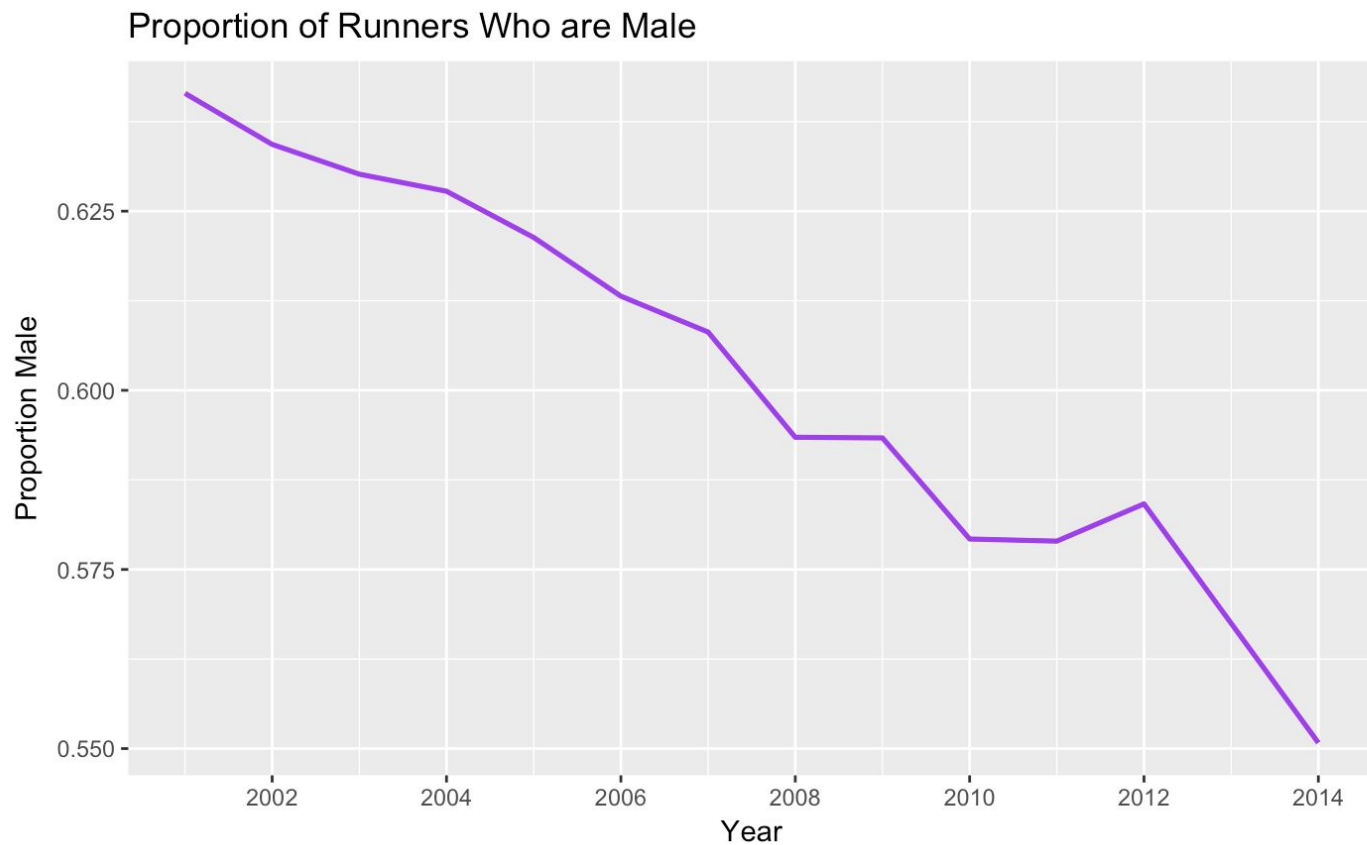
Visualizations (Age)



Visualizations (Gender)

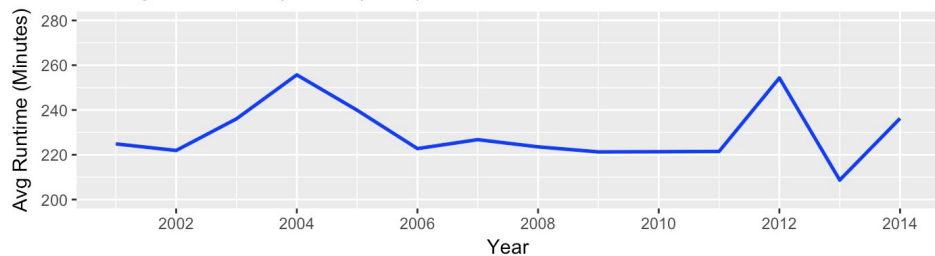


Visualizations (Gender)

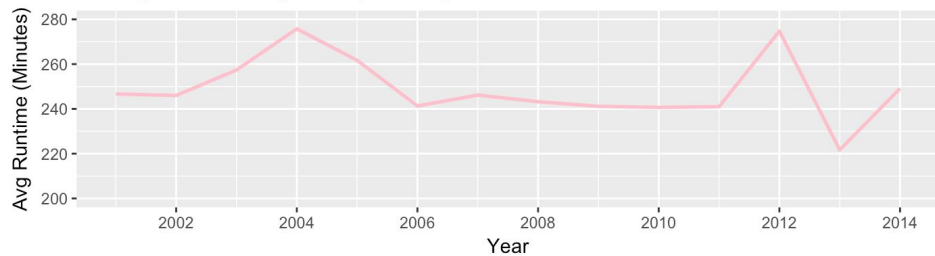


Visualizations (Gender)

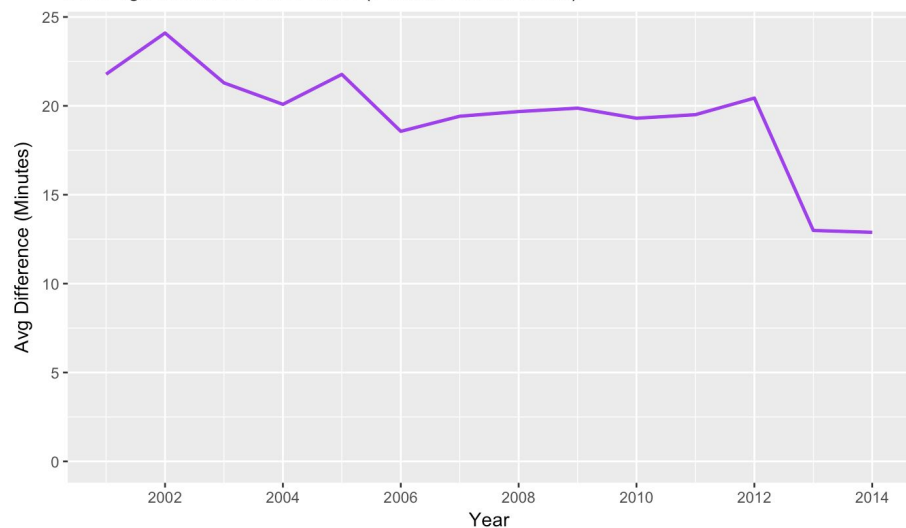
Average Runtime by Year (Male)



Average Runtime by Year (Female)

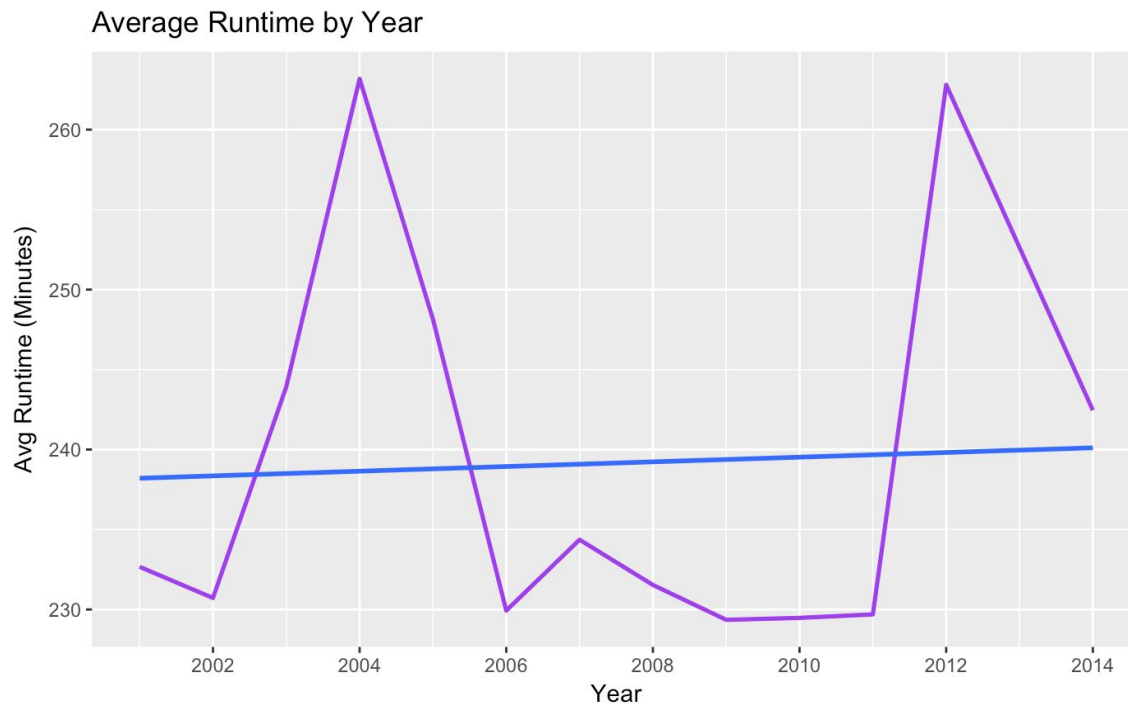


Average Runtime Difference (Female Minus Male)

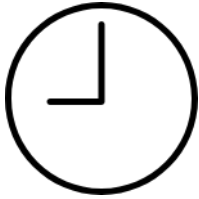


Conclusion and Final Thoughts

Our initial hypothesis is **supported** by data analysis
Average run times increase over time



Challenges and Problems



Format changes
over time



Faulty 2013
dataset



Confounding
factors



Investigate possible
confounding factors



Look for data from
years before 2001

System Description:

- Tableau interactive heat map
- Observing average run time, age, and number of participants by country

<https://public.tableau.com/profile/ash.mohan2079#!/vizhome/bmarathon/Dashboard1>



Appendix

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/

Boston Marathon Runners 2017

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

A Data Analysis of the 2016 Boston Marathon

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

Finishers Boston Marathon 2015, 2016, 2017

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

