CREDIT UNION CHERRY BLOSSOM 10M RUN

Who are the future runners?

DUSTIN BRACY
PAUL HUGGINS
GRACE LANG
BRANUM STEPHAN

TABLE OF CONTENTS

FUTURE RUNNERS

The popularity of the Cherry Blossom run has grown over the years. What can be done to improve the race for our runners to encourage more participation and further growth?

HIGH-LEVEL OBJECTIVES:

Understand the changing profile of runners

…Is more seeding time structure needed?

…Are runners younger/older, faster/slower?

…What is the overall trend, year over year?

MODELING FOR SUCCESS

DELIVERING BUSINESS VALUE

AMONG SEVERAL REASONS, THE FOLLOWING ARE A FEW MAJOR PURPOSES OF PROPER RACE TIMING AND RACER PROFILE UNDERSTANDING

- Safety: it is important to ensure that there aren't too many runners at once in the starting interval in order to ***reduce liability***
- Concessions: Staging can affect flow of runners finishing the race. Improper staging may ***bring revenue to concession stands*** by creating unnecessarily long lines
- Entertainment: Like concessions, timing of live music and performances is important to ensure that ***all participants can enjoy live entertainment***.

DATA EXTRACTION

/ WEB SCRAPE

Leveraged R to pull data iteratively from www.cballtimeresults.org to capture all 14 years of female run result data

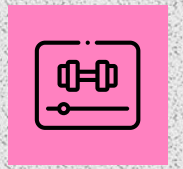
/ NO RECORDS

20 records were removed for missing age or location (less than 0.02% of total data)

/ CLEAN

Converted to proper data types and scrubbed extraneous field strings

/ ENGINEERING



Split columns for further analysis and understanding (e.g. Year, Hometown > City + State)

WEB SCRAPER

Leveraging online data

1999 10M Event Results for Overall Women

Name	Age	Time	Pace	PiS/TiS	Division	PiD/TiD	Hometown
Jane Omoro (W)	26	0:53:37	5:22	1/2358	W2529	1/559	Kenya
Jane Ngotho (W)	29	0:53:38	5:22	2/2358	W2529	2/559	Kenya
Lidiya Grigoryeva (W)	NR	0:53:40	5:22	3/2358	NR	NR	Russia
Eunice Sagero (W)	20	0:53:55	5:24	4/2358	W2024	1/196	Kenya
Alla Zhilyayeva (W)	29	0:54:08	5:25	5/2358	W2529	3/559	Russia
Teresa Wanjiku (W)	24	0:54:10	5:25	6/2358	W2024	2/196	Kenya
Elana Viazova (W)	38	0:54:29	5:27	7/2358	W3539	1/387	Ukraine
Gladys Asiba (W)	NR	0:54:50	5:29	8/2358	NR	NR	Kenya
Nnenna Lynch (W)	27	0:55:39	5:34	9/2358	W2529	4/559	Concord, MA
Margaret Kagiri (W)	30	0:55:43	5:34	10/2358	W3034	1/529	Kenya
Susannah Beck (W)	30	0:56:13	5:37	11/2358	W3034	2/529	Eugene, OR
Kelly Keeler (W)	37	0:57:23	5:44	12/2358	W3539	2/387	Bloomington, MN
Marie Boyd (W)	39	0:57:24	5:44	13/2358	W3539	3/387	Albuquerque, NM
Betsy Kempter (W)	32	0:57:51	5:47	14/2358	W3034	3/529	Chapel Hill, NC
Naoko Ishibe (W)	30	0:58:05	5:49	15/2358	W3034	4/529	Washington, DC
Bea Marie Altieri (W)	31	0:58:36	5:52	16/2358	W3034	5/529	Columbia, MD
Connie Buckwalter (W)	NR	0:59:36	5:58	17/2358	NR	NR	Lancaster, PA
Sharon Servidio (W)	25	0:59:42	5:58	18/2358	W2529	5/559	Alexandria, VA
Donna Moore (W)	38	0:59:48	5:59	19/2358	W3539	4/387	Silver Spring, MD
Marian Huizing (W)	31	1:00:30	6:03	20/2358	W3034	6/529	Rockville, MD

prev next

Sample of page 1 of women's 10k runner results for 1999

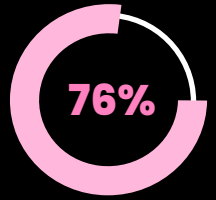
- Tables were parsed from website https://www.cherryblossom.org/
- Tools comprised of R's "tidyverse", "XML2", and more
- Created an algorithm to parse an unknown number of pages for a known time year period
- Performed additional splits and age binning

	Year	Name	Age	Time	Pace	PiS.TiS	Division	PiD.TiD	City	State	Country	TotalTimeSec	TotalTimeMin	PaceTimeSec	PaceTimeMin	AgeGroup
1	1999	Jane Omoro	26	0:53:37	5:22	1/2358	W2529	1/559	<NA>	<NA>	Kenya	3217	53.617	322	5.367	20s
2	1999	Jane Ngotho	29	0:53:38	5:22	2/2358	W2529	2/559	<NA>	<NA>	Kenya	3218	53.633	322	5.367	20s
3	1999	Lidiya Grigoryeva	NA	0:53:40	5:22	3/2358	<NA>	<NA>	<NA>	<NA>	Russia	3220	53.667	322	5.367	<NA>
4	1999	Eunice Sagero	20	0:53:55	5:24	4/2358	W2024	1/196	<NA>	<NA>	Kenya	3235	53.917	324	5.400	20s
5	1999	Alla Zhilyayeva	29	0:54:08	5:25	5/2358	W2529	3/559	<NA>	<NA>	Russia	3248	54.133	325	5.417	20s
6	1999	Teresa Wanjiku	24	0:54:10	5:25	6/2358	W2024	2/196	<NA>	<NA>	Kenya	3250	54.167	325	5.417	20s
7	1999	Elana Viazova	38	0:54:29	5:27	7/2358	W3539	1/387	<NA>	<NA>	Ukraine	3269	54.483	327	5.450	30s
8	1999	Gladys Asiba	NA	0:54:50	5:29	8/2358	<NA>	<NA>	<NA>	<NA>	Kenya	3290	54.833	329	5.483	<NA>
9	1999	Nnenna Lynch	27	0:55:39	5:34	9/2358	W2529	4/559	Concord	MA	USA	3339	55.650	334	5.567	20s
10	1999	Margaret Kagiri	30	0:55:43	5:34	10/2358	W3034	1/529	<NA>	<NA>	Kenya	3343	55.717	334	5.567	30s
11	1999	Susannah Beck	30	0:56:13	5:37	11/2358	W3034	2/529	Eugene	OR	USA	3373	56.217	337	5.617	30s
12	1999	Kelly Keeler	37	0:57:23	5:44	12/2358	W3539	2/387	Bloomington	MN	USA	3443	57.383	344	5.733	30s
13	1999	Marie Boyd	39	0:57:24	5:44	13/2358	W3539	3/387	Albuquerque	NM	USA	3444	57.400	344	5.733	30s
14	1999	Betsy Kempter	32	0:57:51	5:47	14/2358	W3034	3/529	Chapel Hill	NC	USA	3471	57.850	347	5.783	30s
15	1999	Naoko Ishibe	30	0:58:05	5:49	15/2358	W3034	4/529	Washington	DC	USA	3485	58.083	349	5.817	30s
16	1999	Bea Marie Altieri	31	0:58:36	5:52	16/2358	W3034	5/529	Columbia	MD	USA	3516	58.600	352	5.867	30s
17	1999	Connie Buckwalter	NA	0:59:36	5:58	17/2358	<NA>	<NA>	Lancaster	PA	USA	3576	59.600	358	5.967	<NA>
18	1999	Sharon Servidio	25	0:59:42	5:58	18/2358	W2529	5/559	Alexandria	VA	USA	3582	59.700	358	5.967	20s
19	1999	Donna Moore	38	0:59:48	5:59	19/2358	W3539	4/387	Silver Spring	MD	USA	3588	59.800	359	5.983	30s
20	1999	Marian Huizing	31	1:00:30	6:03	20/2358	W3034	6/529	Rockville	MD	USA	3630	60.500	363	6.050	30s
21	1999	Martha Merz	36	1:00:32	6:03	21/2358	W3539	5/387	Annandale	VA	USA	3632	60.533	363	6.050	30s
22	1999	Wendy Nelson-Barrett	30	1:00:43	6:04	22/2358	W3034	7/529	Lebanon	PA	USA	3643	60.717	364	6.067	30s
23	1999	Patti Shull	40	1:00:47	6:05	23/2358	W4044	1/306	Ashburn	VA	USA	3647	60.783	365	6.083	40s
24	1999	Anita Freres	34	1:01:04	6:06	24/2358	W3034	8/529	Rockville	MD	USA	3664	61.067	366	6.100	30s
25	1999	Aryn Fahey	22	1:01:15	6:08	25/2358	W2024	3/196	Fairfax	VA	USA	3675	61.250	368	6.133	20s
26	1999	Desiree Ficker	22	1:01:20	6:08	26/2358	W2024	4/196	Potomac	MD	USA	3680	61.333	368	6.133	20s
27	1999	Kate Petricone	37	1:01:59	6:12	27/2358	W3539	6/387	Winsted	CT	USA	3719	61.983	372	6.200	30s
28	1999	Jennifer Janis	29	1:02:07	6:13	28/2358	W2529	6/559	Malvern	PA	USA	3727	62.117	373	6.217	20s
29	1999	Christina Heming	34	1:02:12	6:13	29/2358	W3034	9/529	Shorewood	WI	USA	3732	62.200	373	6.217	30s
30	1999	Patty Fulton	33	1:02:16	6:14	30/2358	W3034	10/529	Silver Spring	MD	USA	3736	62.267	374	6.233	30s
31	1999	Nancy Watkins	30	1:02:28	6:15	31/2358	W3034	11/529	Stafford	VA	USA	3748	62.467	375	6.250	30s
32	1999	Kirsten Black	24	1:02:45	6:17	32/2358	W2024	5/196	Arlington	VA	USA	3765	62.750	377	6.283	20s
33	1999	Dorian Meyer	39	1:02:51	6:17	33/2358	W3539	7/387	Rumson	NJ	USA	3771	62.850	377	6.283	30s
34	1999	Karen Oudekerk	31	1:03:01	6:18	34/2358	W3034	12/529	Arlington	VA	USA	3781	63.017	378	6.300	30s
35	1999	Anne Britt	34	1:03:10	6:19	35/2358	W3034	13/529	Summit	NJ	USA	3790	63.167	379	6.317	30s
36	1999	Denise Knickman	30	1:03:31	6:21	36/2358	W3034	14/529	Baltimore	MD	USA	3811	63.517	381	6.350	30s
37	1999	Cecily Tynan	30	1:03:33	6:21	37/2358	W3034	15/529	Philadelphia	PA	USA	3813	63.550	381	6.350	30s
38	1999	Leslie Minnix-Wolfe	37	1:03:53	6:23	38/2358	W3539	8/387	Reston	VA	USA	3833	63.883	383	6.383	30s
39	1999	Meg Ritter	25	1:03:56	6:24	39/2358	W2529	7/559	Richmond	VA	USA	3836	63.933	384	6.400	20s
40	1999	Debi Bernardes	40	1:04:15	6:26	40/2358	W4044	2/306	King George	VA	USA	3855	64.250	386	6.433	40s
41	1999	Linda Wack	43	1:05:01	6:30	41/2358	W4044	3/306	Germantown	MD	USA	3901	65.017	390	6.500	40s
42	1999	Patricia Keating	37	1:05:17	6:32	42/2358	W3539	9/387	Clarksville	MD	USA	3917	65.283	392	6.533	30s

Parsed table of women's run data, holding more than 75,000 records

DATA

/ AGE



Women in their 20's & 30's

/ TOP STATES

VA, MA & DC

Runners come from just 3 states

/ AVERAGE TIME

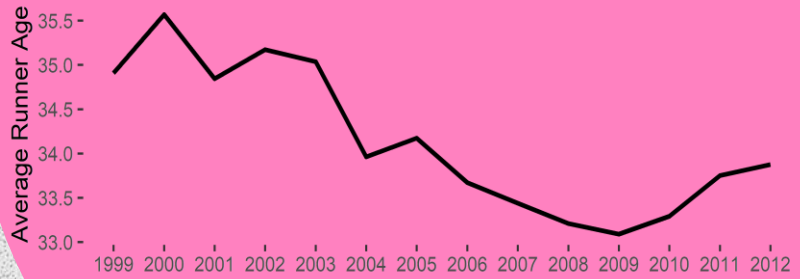
01:39:01

*Data from 2012 run

CHERRY BLOSSOM
WOMEN'S 10M RUN

RUNNER TRENDS

The average runner age has decreased since inception of race in 1999



+0.29%
Average Race Time Trend (1999-2012)

12%

The average growth of number of runners each year

RACER SPOTLIGHT

Muffet Chatterton
Crofton, MD

Year	Age	Pace	PiD.TiD
1999	54	8:10	10/114
2000	55	8:13	6/60
2001	56	8:26	10/66
2002	57	9:06	17/72
2003	58	8:43	16/83
2004	59	9:13	17/83
2006	61	8:46	8/48
2007	62	9:22	5/33
2008	63	9:34	19/50
2009	64	9:43	17/56
2010	65	9:57	6/17
2011	66	9:52	12/26
2012	67	10:07	14/35

PiD = Position in Division
TiD = Total in Division

Muffet has ran the Cherry Blossom race from 1999-2012. She was unable to attend in 2005 but came back in 2006 with a vengeance posting a blistering 8:46 pace!

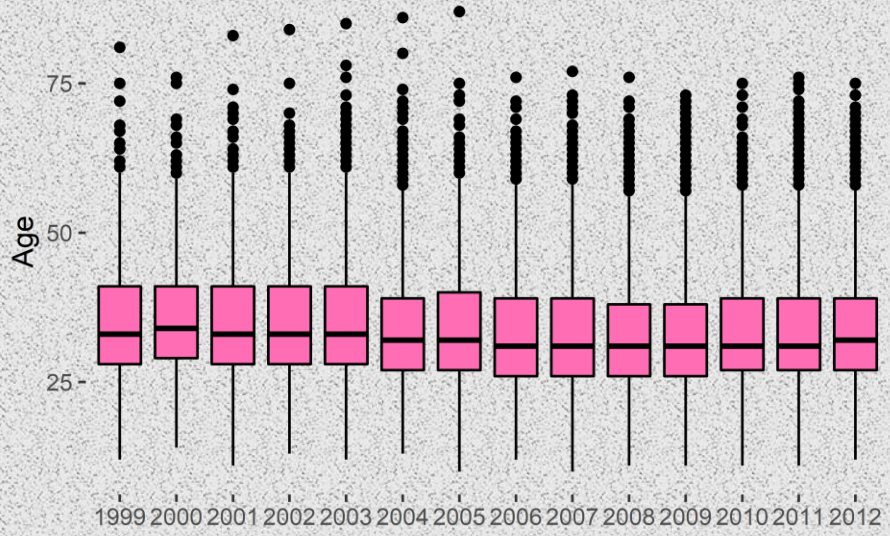
She routinely finished in the top 15 in her division year after year and completed the 2012 race at age 67.

Cheers to you, Muffet!

FAST FACTS

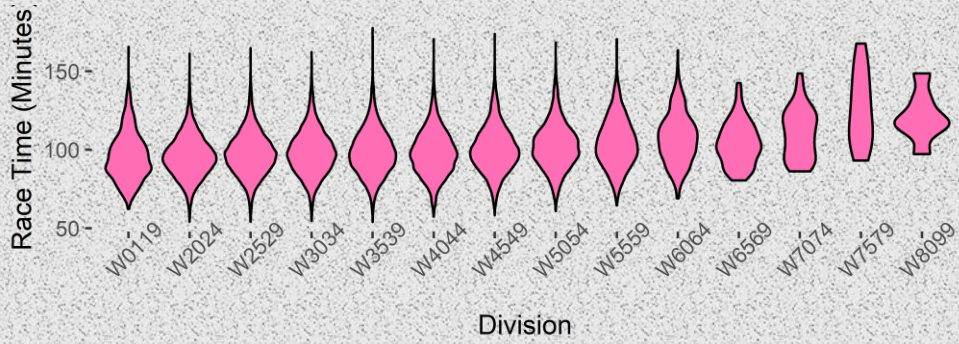
Runner Age Distribution

Runner age distribution has stayed relatively consistent year over year.



Race Time by Division

W3539 sees the widest range of finish times compared to W8099 with the smallest range.



ANOVA METHOD

Analysis of variance, or ANOVA, is a statistical method used to compare averages of more than two groups.

In this analysis, ANOVA is used to compare average race time in minutes across age groups and/or by year.

An ANOVA that has a significant p-value (< 0.05), or a confidence interval that does not include zero, suggests that the averages in comparison are statistically different than one another.

ANOVA RESULTS

AGE COMPARISON	AVG TIME DIFF	LOWER CONF INT	UPPER CONF INT	P-VALUE * Tukey Adjusted
01-19/20-29	1.1441	-0.5264	2.8146	**0.4307**
70-79/80-89	7.0590	-9.8772	23.9952	**0.9123**
60-69/80-89	14.4883	-1.6633	30.6399	**0.1168**

The above chart highlights where p-values are **not significant**, indicating there is **no significant difference in mean run times between age groups.** There is a significant difference in run times between every other age combination outside of this table. At a 95% confidence level, the upper and lower bounds include zero, which is another indication that zero is a plausible difference in mean run time for the groups above.

RECOMMENDATIONS

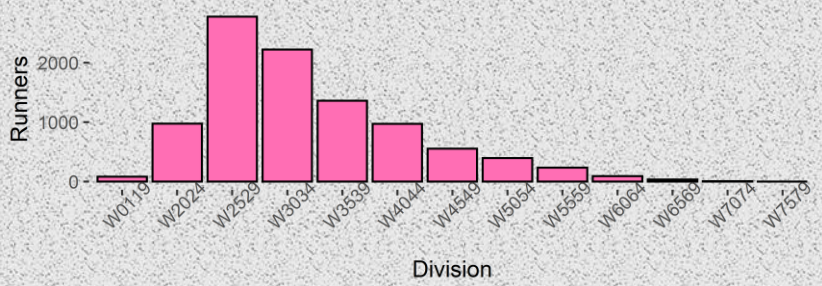
10M RUN – OPERATIONS DAY OF

- Start slow divisions earlier
- Expect approximately 10,000 runners for 2013
- Break larger divisions into multiple gates

NEW DIVISIONS

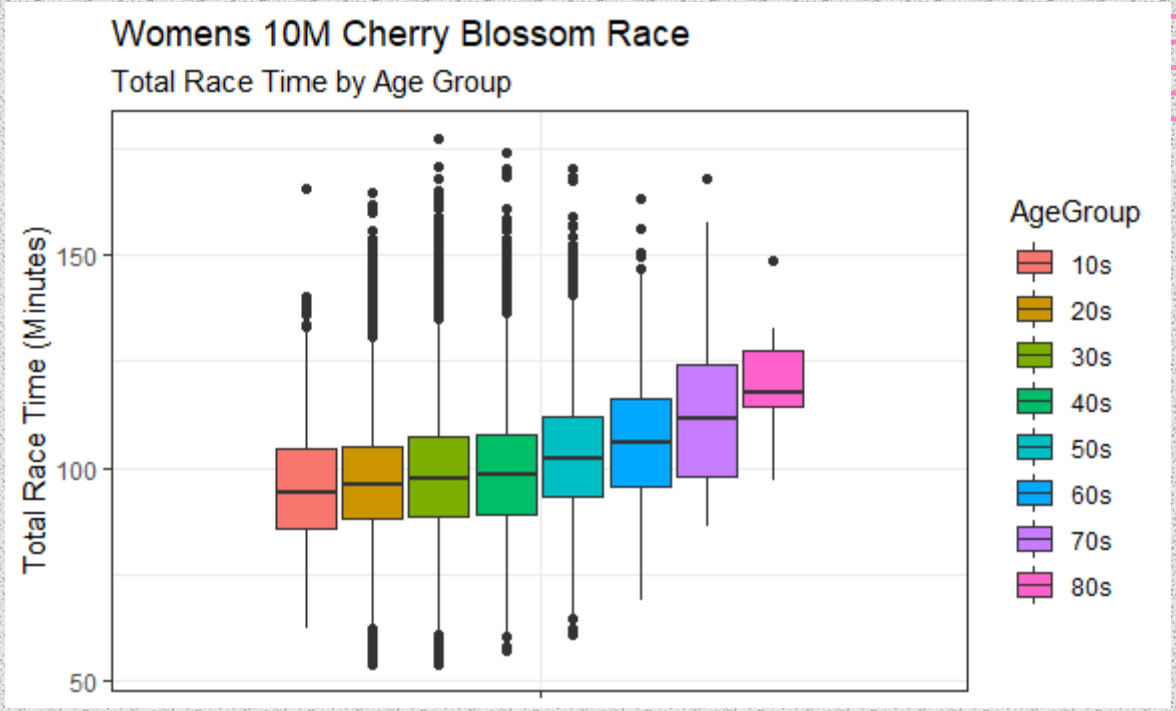
- Under 30 Club (combines 0-19, 20-24, 25-29)
  - Largest runner population in this group
  - No significant difference in time
- Golden Years (combines 70-74, 75-79, 80-100)

NUMBER OF RUNNERS BY DIVISION



APPENDIX

AVERAGE RACE TIME VARIES BY AGE GROUP



AVERAGE RACE TIME WITHIN AN AGE GROUP

VARIES VERY LITTLE