Assignment crt2021 dublin marathon visualization in R

Aryan Singh

9/28/2021

Contents

Introduction

This is a submission for assignment **Visualization in R using ggplot2** Link to assignment

The content covers:

- Marathon data set
- Questions that can be answered using the data set
- Conclusion

The Marathon data set

```
rows <- nrow(df)
cols <- ncol(df)

datatable(df)
```

```
## Warning in instance$preRenderHook(instance): It seems your data is too big
## for client-side DataTables. You may consider server-side processing: https://
## rstudio.github.io/DT/server.html
```

Show 10 entries Search:

	...1	Race Number	First Name	Surname	Gender	Gender Position	Category	Category Position	Club	10K	Stage Position...11	Halfway	Stage Position...13	30K	Stage Position...15	Gun Time	Overall Position	Chip Time	Chip Position	YouTube	Your Race Video	Share
1	1	2	Asefa	Bekele	Male	1	MS	1		00:32:12	9	01:06:50	4	01:34:46	4	02:13:24	1	02:13:23	1		Race Video	
2	2	13	David	Manja	Male	2	MS	2		00:32:12	3	01:06:49	3	01:34:45	1	02:13:33	2	02:13:33	2		Race Video	
3	3	5	Joel	Kiptoo	Male	3	MS	3		00:32:11	1	01:06:50	5	01:34:45	2	02:13:42	3	02:13:41	3		Race Video	
4	4	1	Benard	Rotich	Male	4	MS	4		00:32:12	4	01:06:51	8	01:34:48	7	02:14:18	4	02:14:18	4		Race Video	
5	5	24	Seth	Totten	Male	5	MS	5		00:32:13	12	01:06:50	6	01:34:47	5	02:14:30	5	02:14:28	5		Race Video	
6	6	28	Mick	Clohisey	Male	6	MS	6	RAHENY SHAMROCK A.C.	00:32:13	13	01:07:10	16	01:36:13	13	02:15:58	6	02:15:57	6		Race Video	
7	7	23	Josh	Griffiths	Male	7	MS	7		00:32:23	19	01:08:21	20	01:36:53	14	02:16:09	7	02:16:09	7		Race Video	
8	8	21	Fedor	Shutov	Male	8	MS	8		00:32:56	24	01:09:28	27	01:37:53	17	02:16:48	8	02:16:47	8		Race Video	
9	9	18	Tadses	Tola	Male	9	MS	9		00:32:12	5	01:06:50	7	01:34:47	6	02:17:03	9	02:17:03	9		Race Video	
10	10	26	Gary	O Hanlon	Male	10	M40	1	CLONLIFFE HARRIERS A.C.	00:32:22	18	01:08:16	19	01:37:09	15	02:17:11	10	02:17:10	10		Race Video	

Showing 1 to 10 of 16,433 entries Previous 1 2 3 4 5 ... 1644 Next

Data frame has 22 variables/columns and 16433 measurement/rows in marathon data set

Some questions appropriate for the data:

- Which age group participates actively in marathons?
- How the distribution looks like for finishers (gender wise) ?
- What was the percentage of exit/disqualification among participants (age wise) ?
- How reliable is the gun time measurement in long races like marathon ?
- What are different types of participants (runner/jogger/walker) ?
- Is there a strong correlation among gender,category,time taken between 1st and 2nd stage , 2nd stage to halfway and halfway to finish ?
- Among top 10 finishers , How were their ranks over different stages of race ?

figure below shows count of participants age wise :

it can be said the most of the participants are from young and middle age and the pattern is similar for both gender.

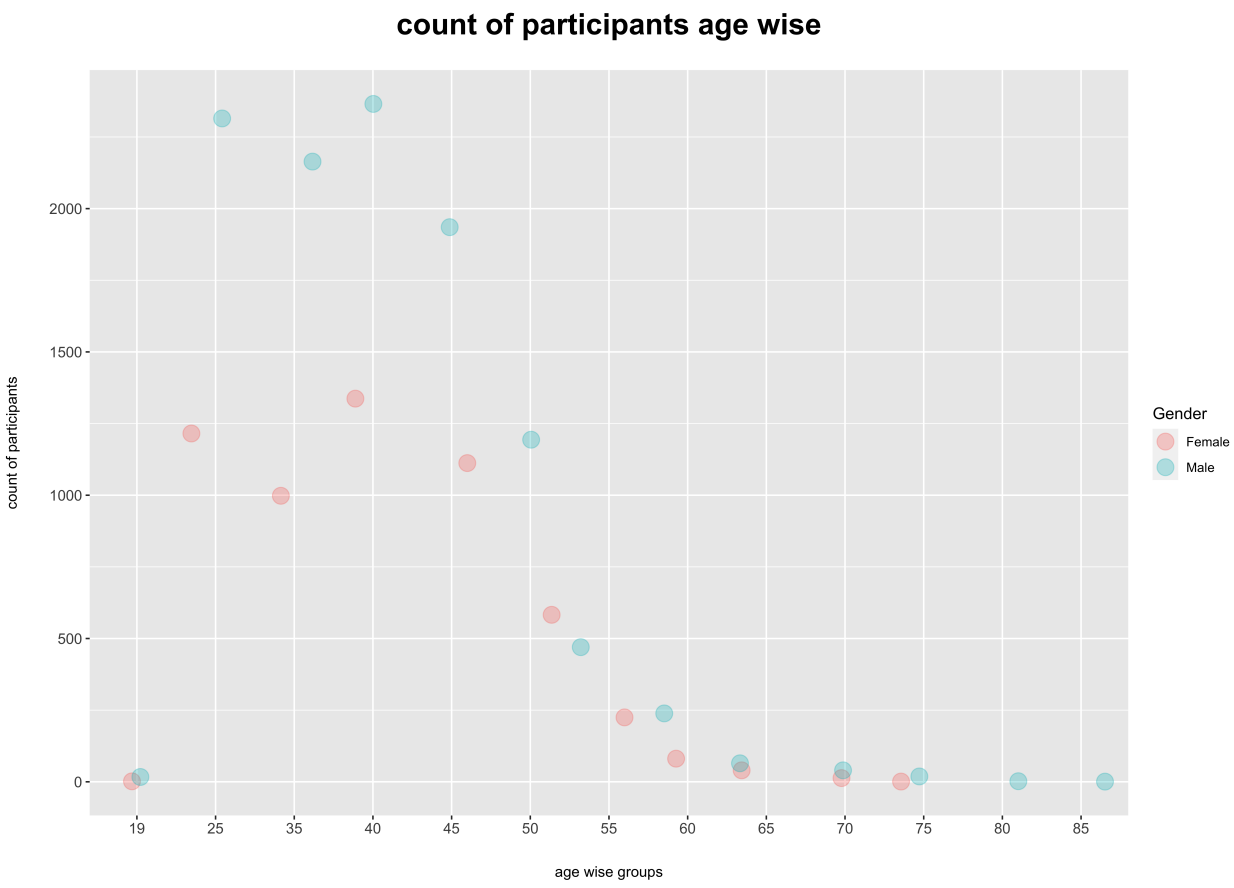


figure below shows the frequency of time at which participants have finished the race

it can be seen from the below diagram that men have finished the race faster than women, but the overall distribution of data remains the same which can imply that both gender are competing optimally given that number of female participants is around 5607 while men participants is around 10826.

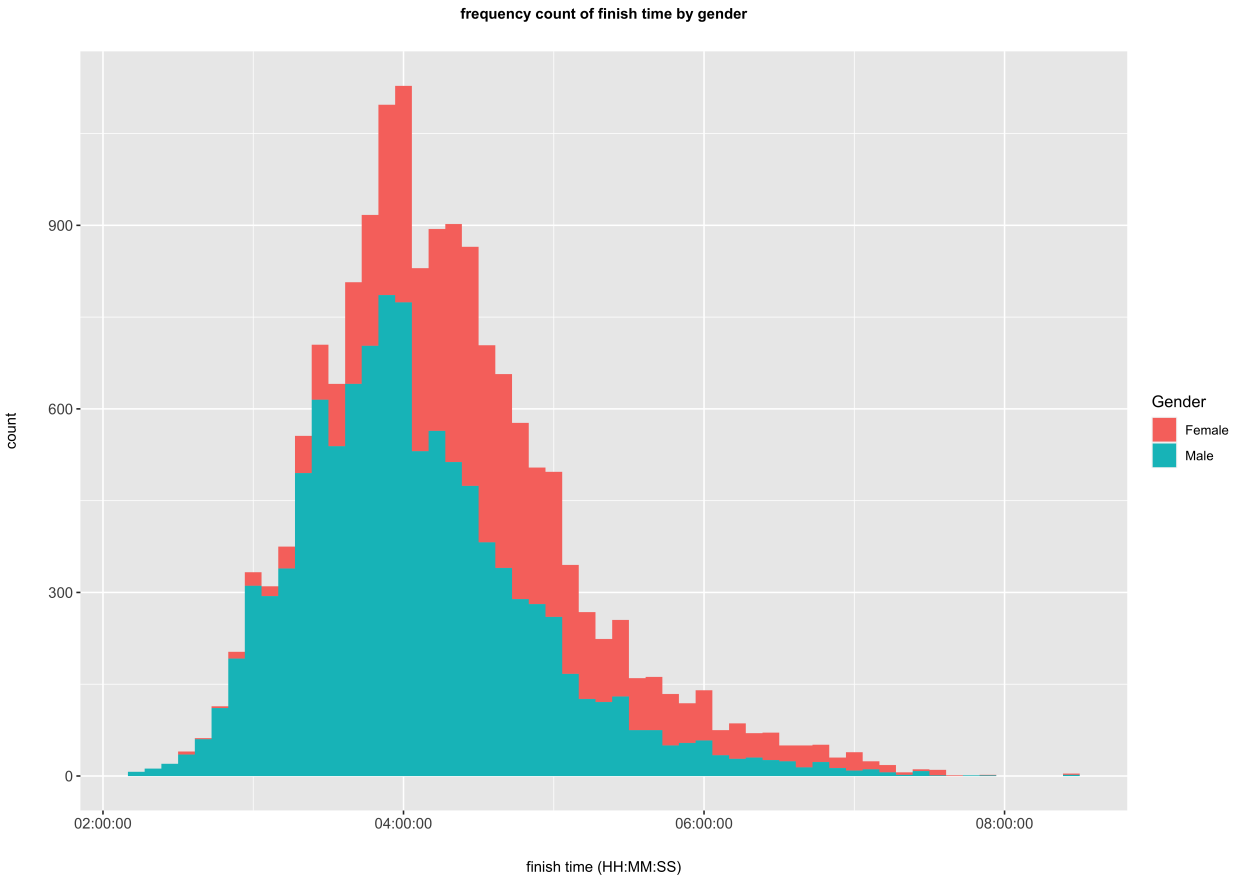


figure below shows the percentage of disqualification / incompletion of race in various age groups and gender

it can be seen that older women have slightly more disqualification as compared to men of the same age group

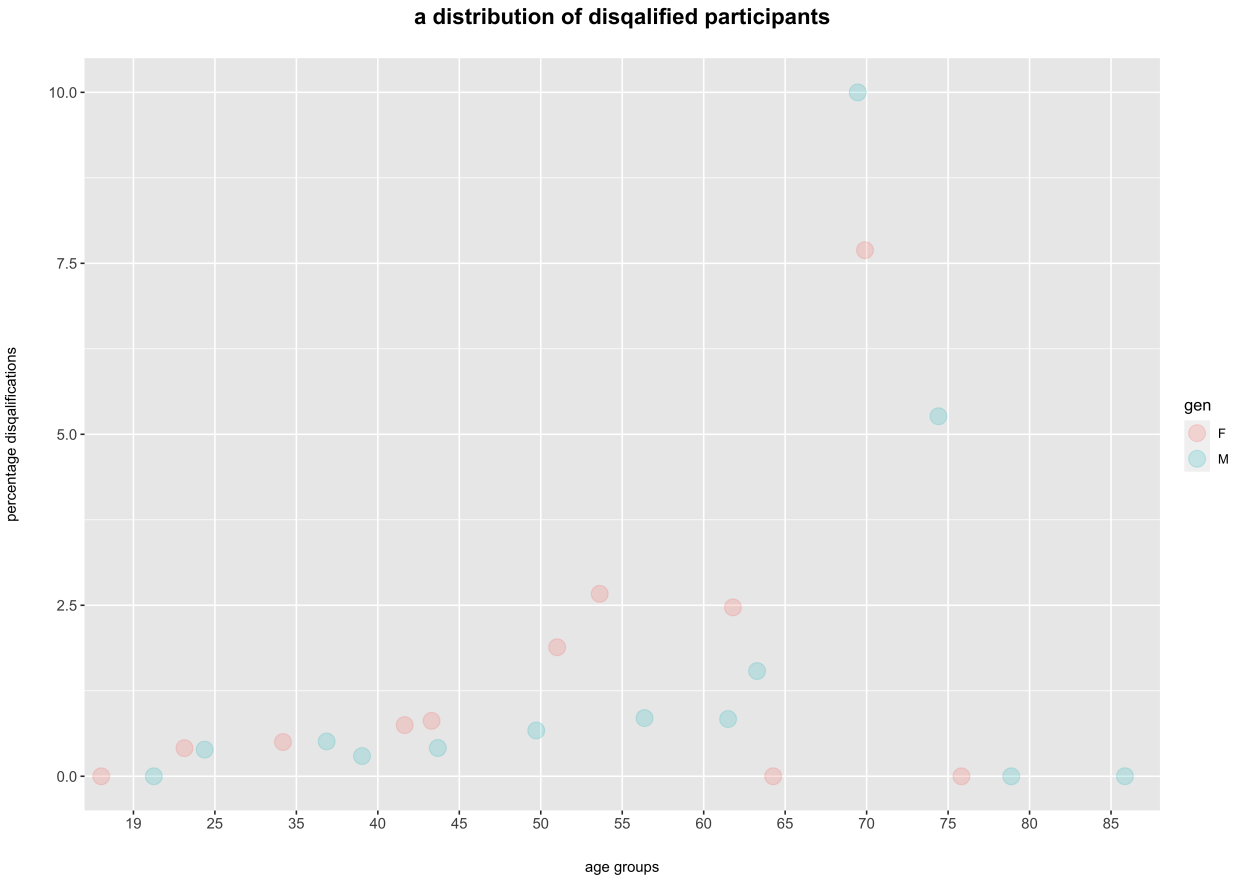


figure below shows a comparision between chip time and gun time.

it can be seen from the plot that gun time is not a reliable rather a ceremonial way to measure finish time , as it deviates a-lot from the actual chip time

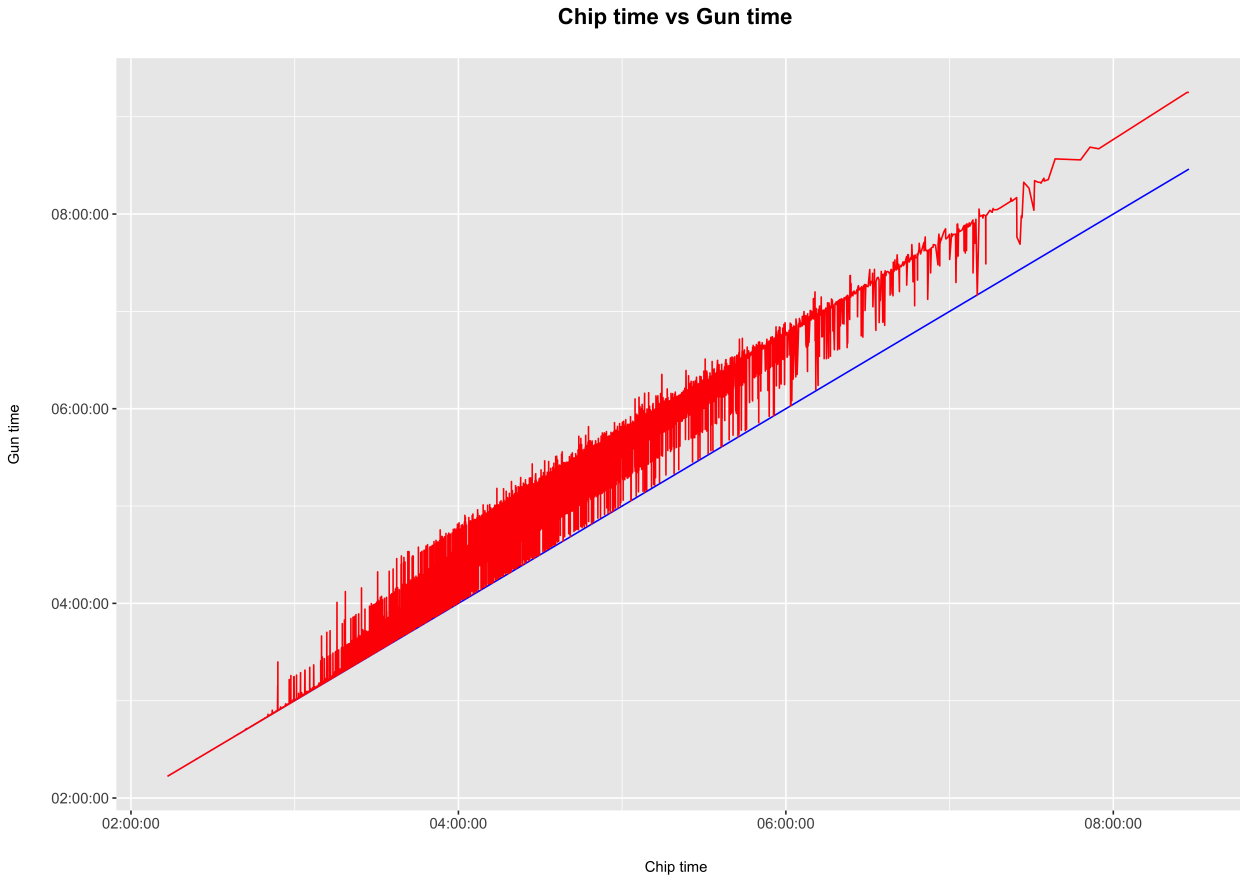


figure below shows type of participants , type can be runner, jogger and walker

based on race finishing time , participants are runner if the finishing time is below 3hrs , jogger if finishing time is between 3hrs and 5hrs and walkers if finishing time is more than 5hrs

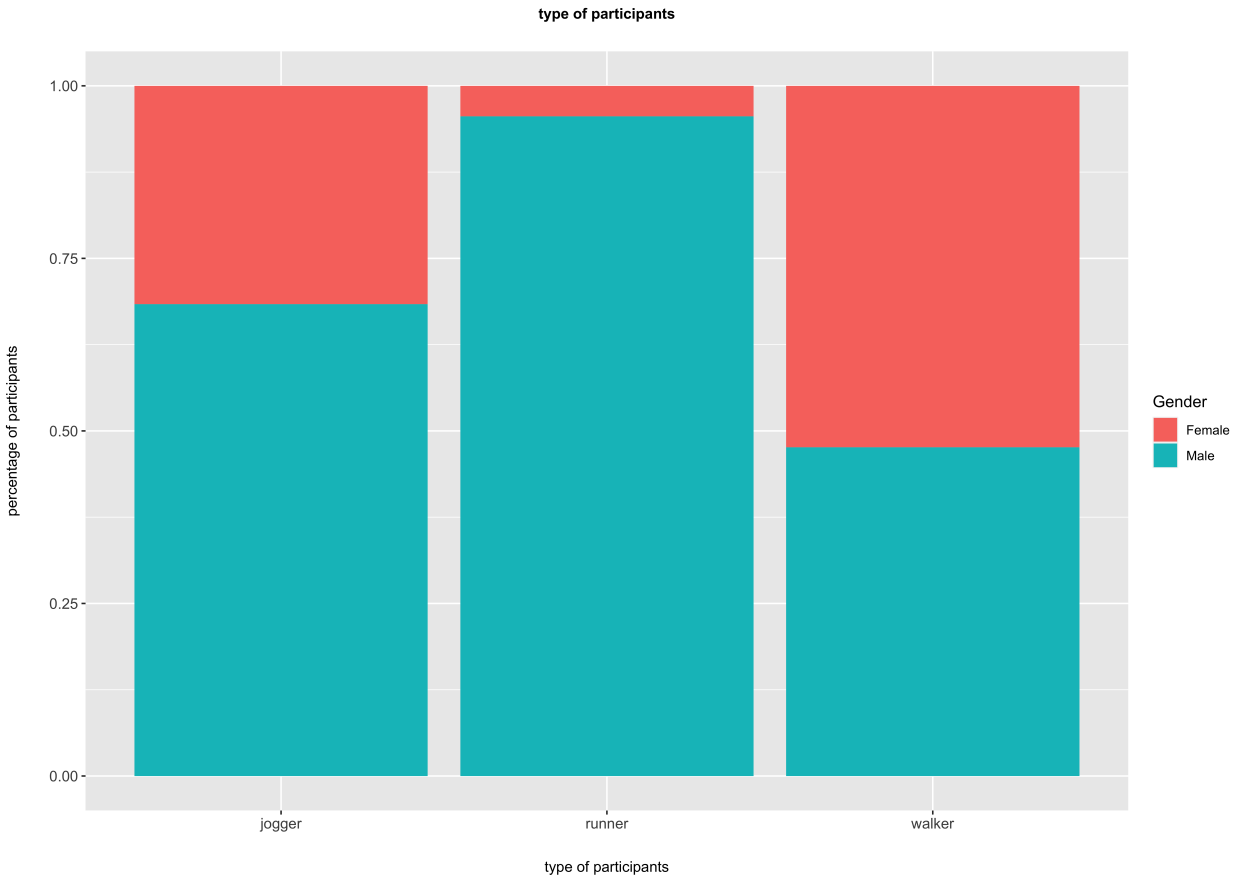


figure below shows correlation between vaiables like gender, category, time taken at various stages of marathon

conclusions can be drawn that, fast finishers at early stages are more likely to have better overall position, which is obvious

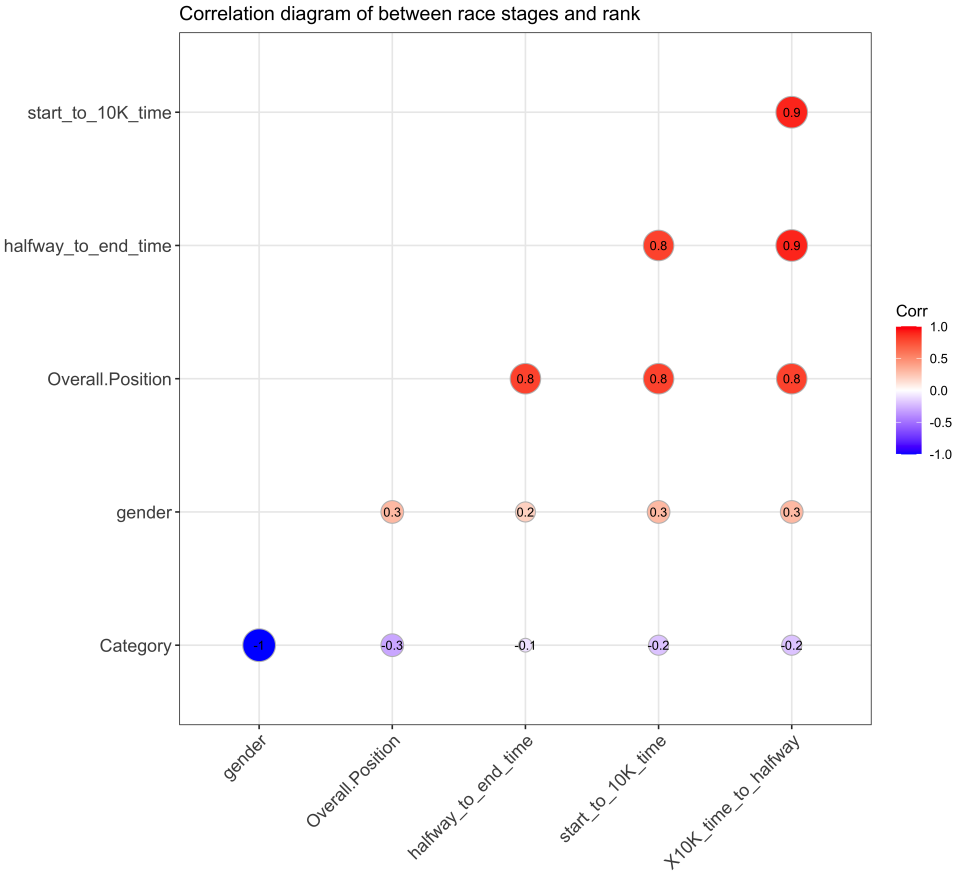
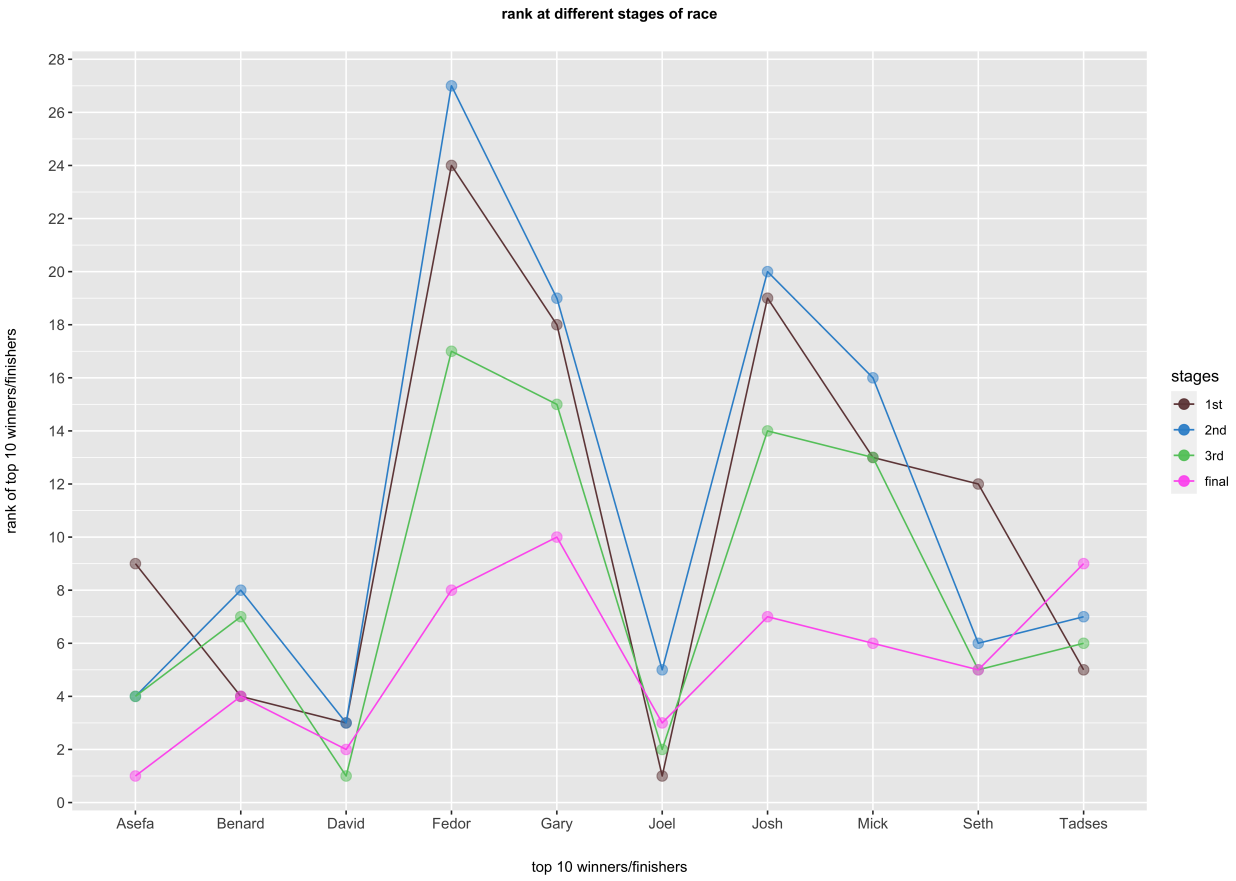


figure below shows positons of top 10 finishers at different stages of marathon

it can be seen that David (2nd position) actually performed better through out the race except for the last stage. He has a high chance of winning any future marathon as his performance is consistent



Link to code and files