assignment_crt2021_dublin_marathon_viz

aryan singh

9/28/2021

Contents

Introduction **1**

The Marathon data set **1**

Some questions appropriate for the data: **2**

figure below shows count of participants age wise : 2

figure below shows the frequency of time at which participants has finished the race 2

figure below shows the percentage of disqalification / incompletion of race in various age groups and gender 3

figure below shows a comparision between chip time and gun time. 4

figure below shows type of participants , type can be runner, jogger and walker 5

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)
```

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

Some questions appropriate for the data:

- Count of participants age wise , 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 between gender,category,time taken between 1st and 2nd stage , 2nd stage to halfway and halfway to finish ?
- Among top 10 finishers , How was their rank over different stages of race ?

figure below shows count of participants age wise :

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

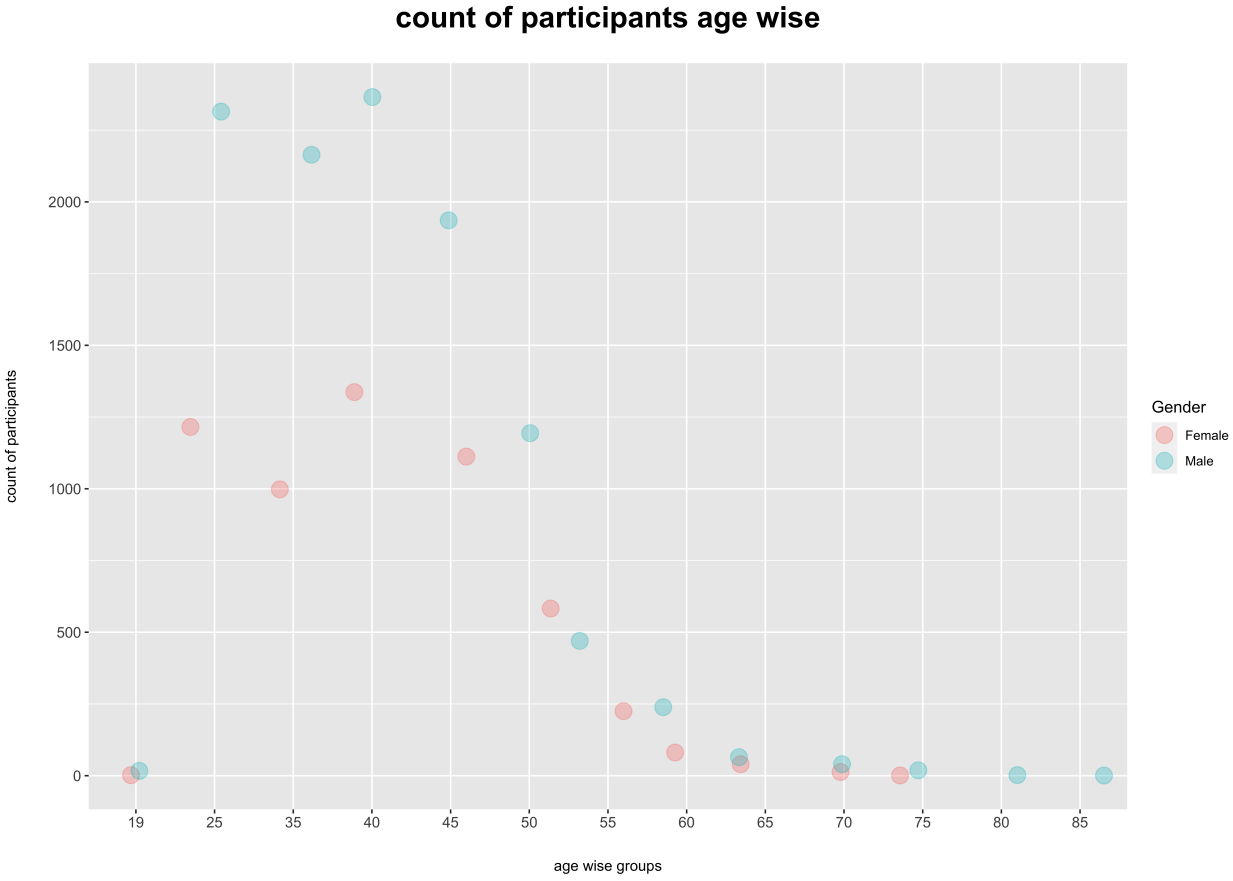


figure below shows the frequency of time at which participants has finished the race

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

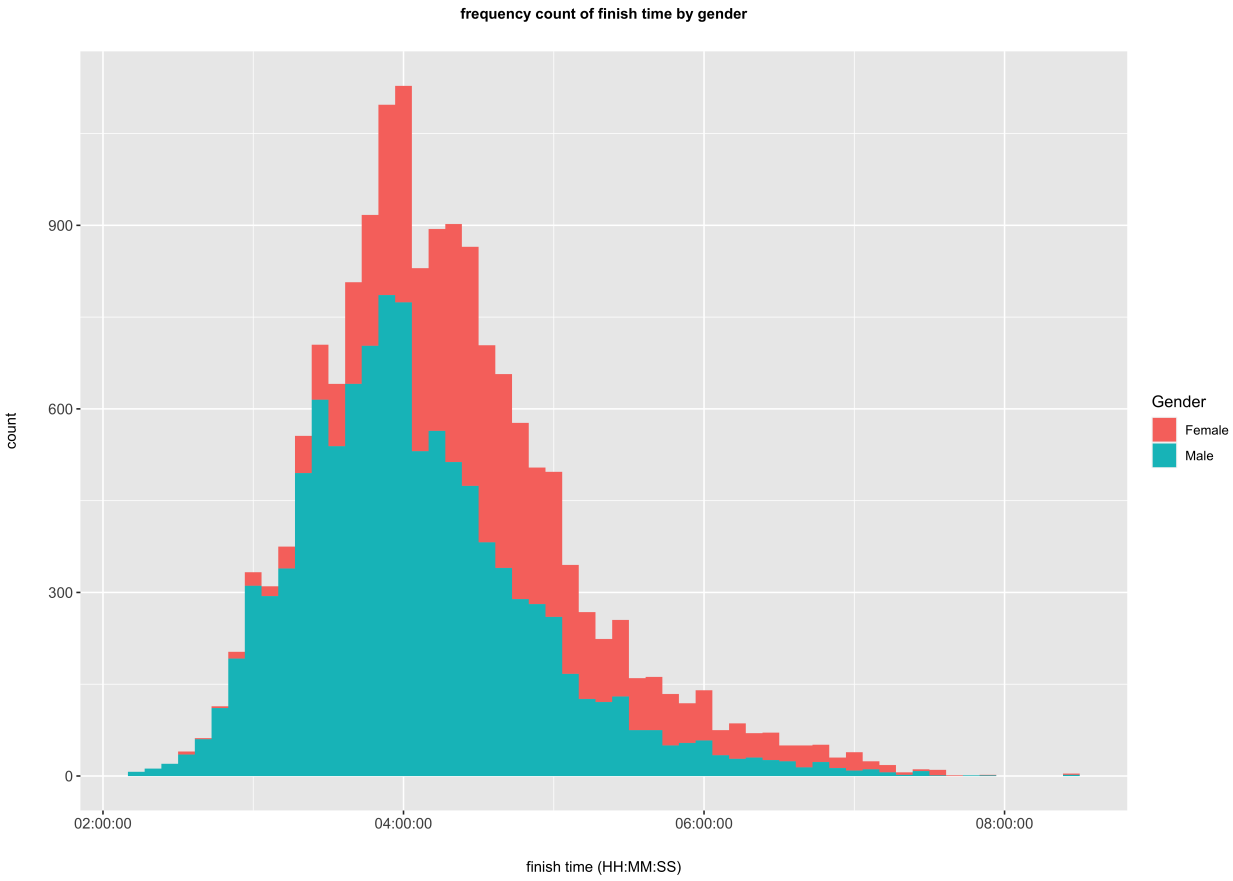


figure below shows the percentage of disqalification / incompletion of race in various age groups and gender

it can be seen that women have a light more disqualification in old ages than men

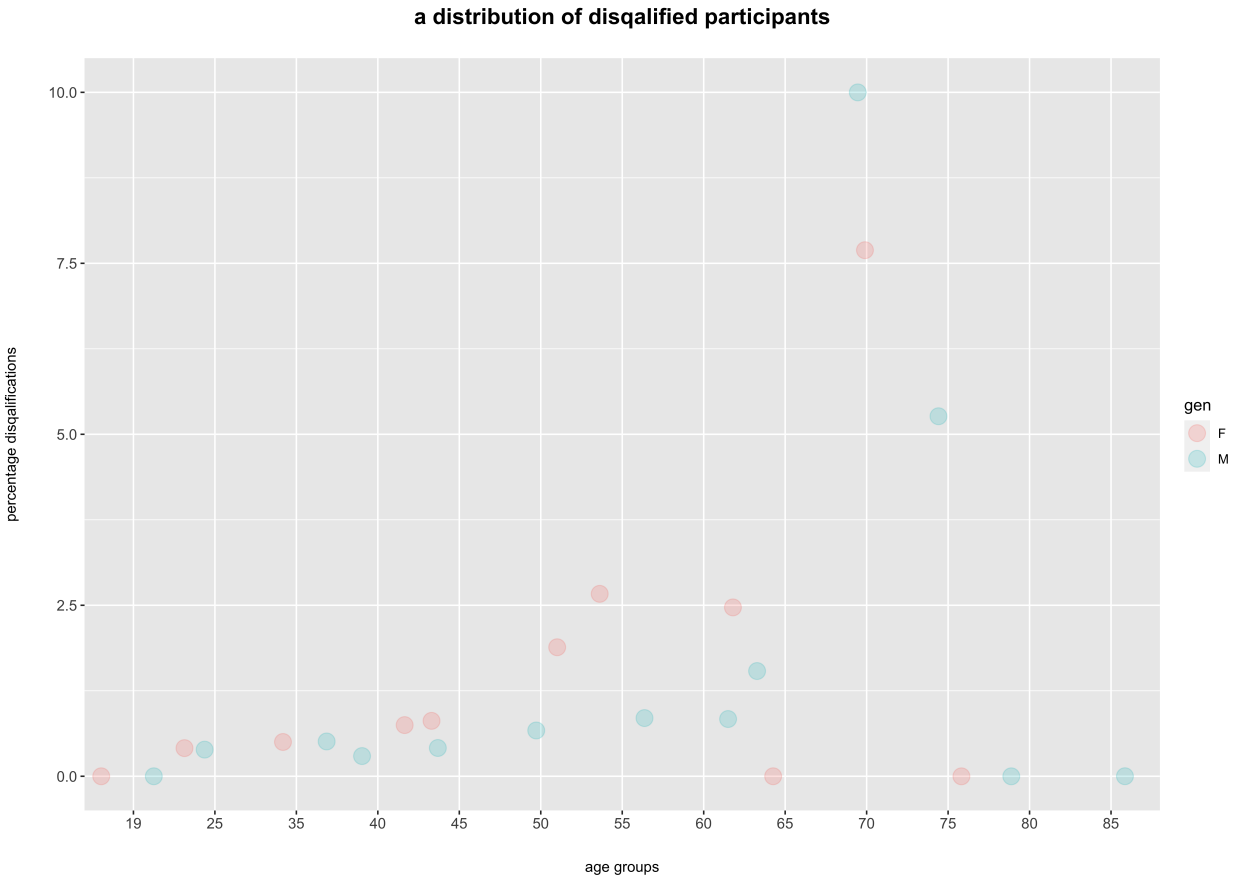


figure below shows a comparision between chip time and gun time.

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

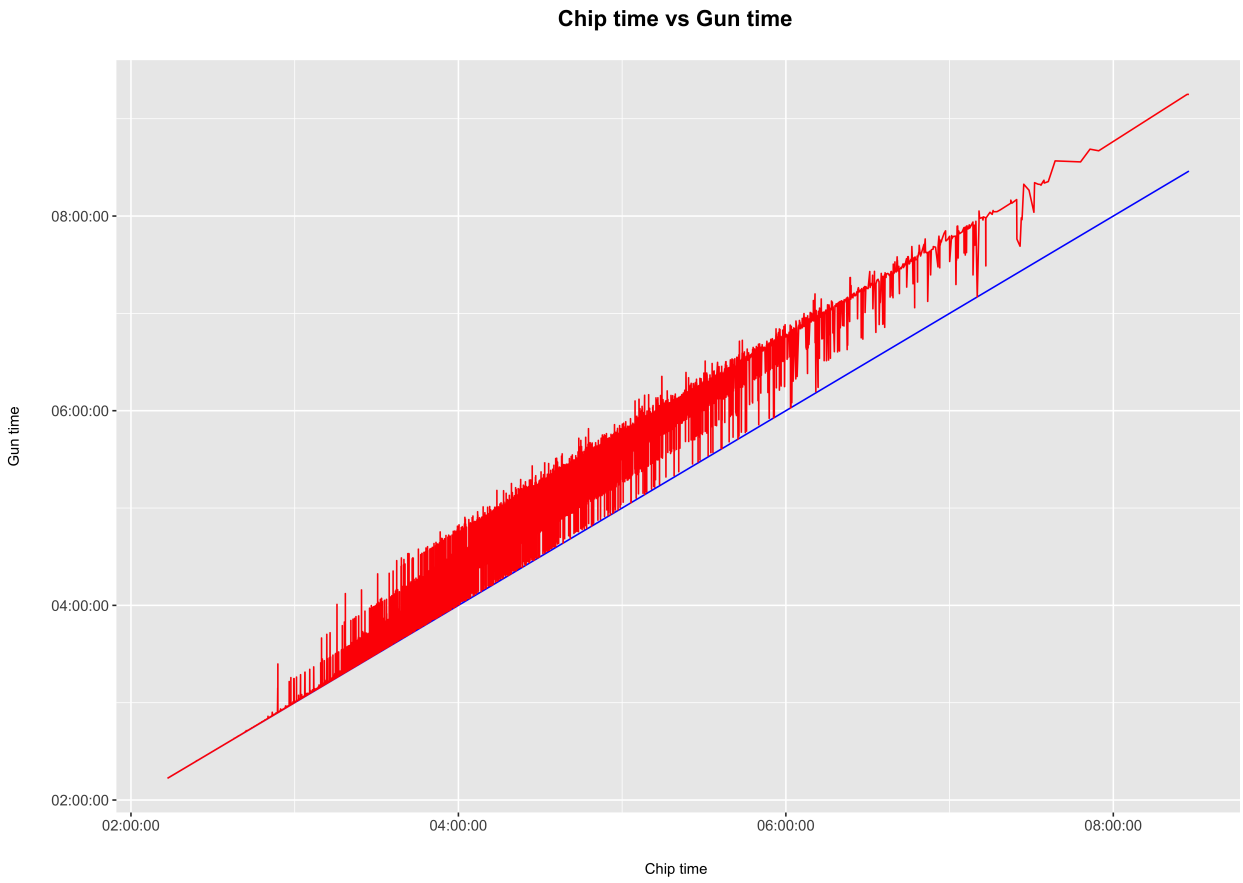
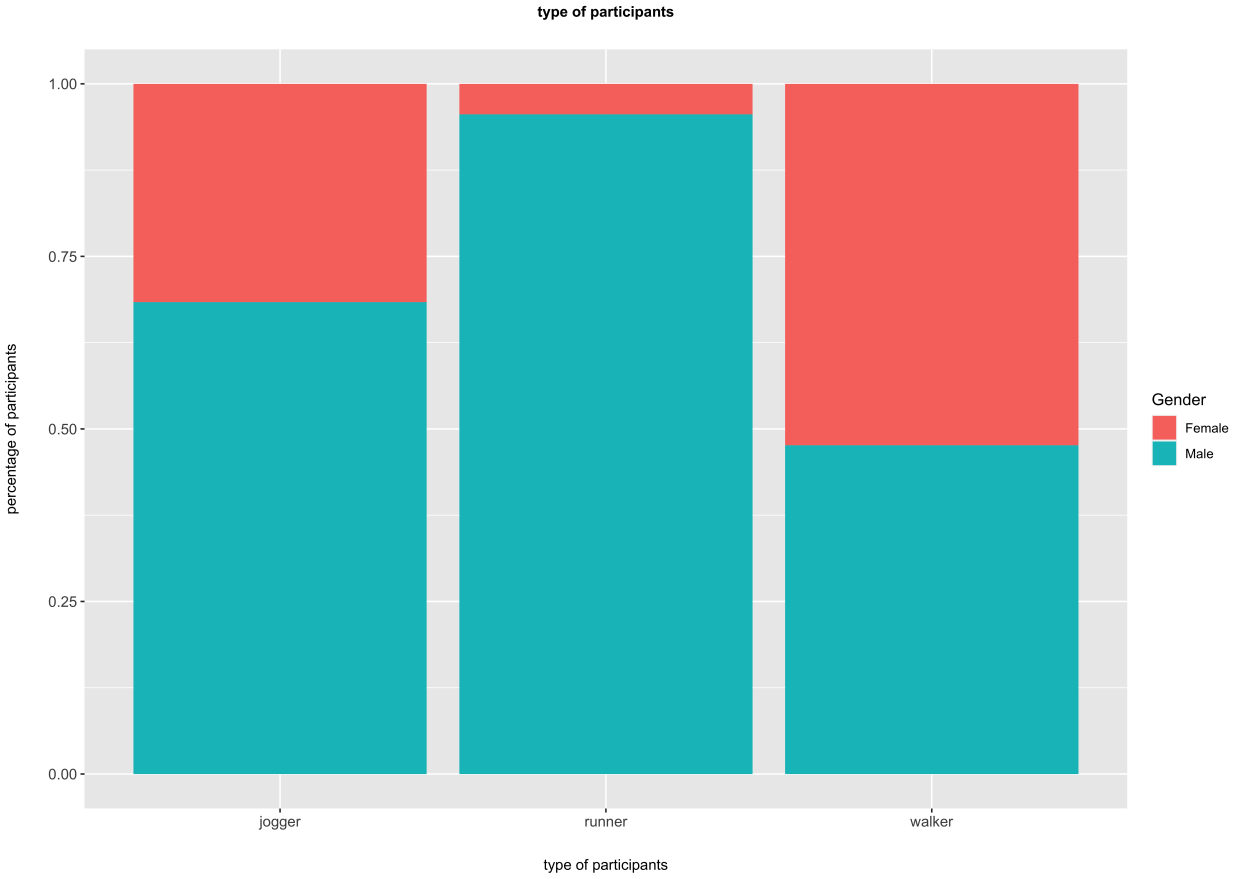
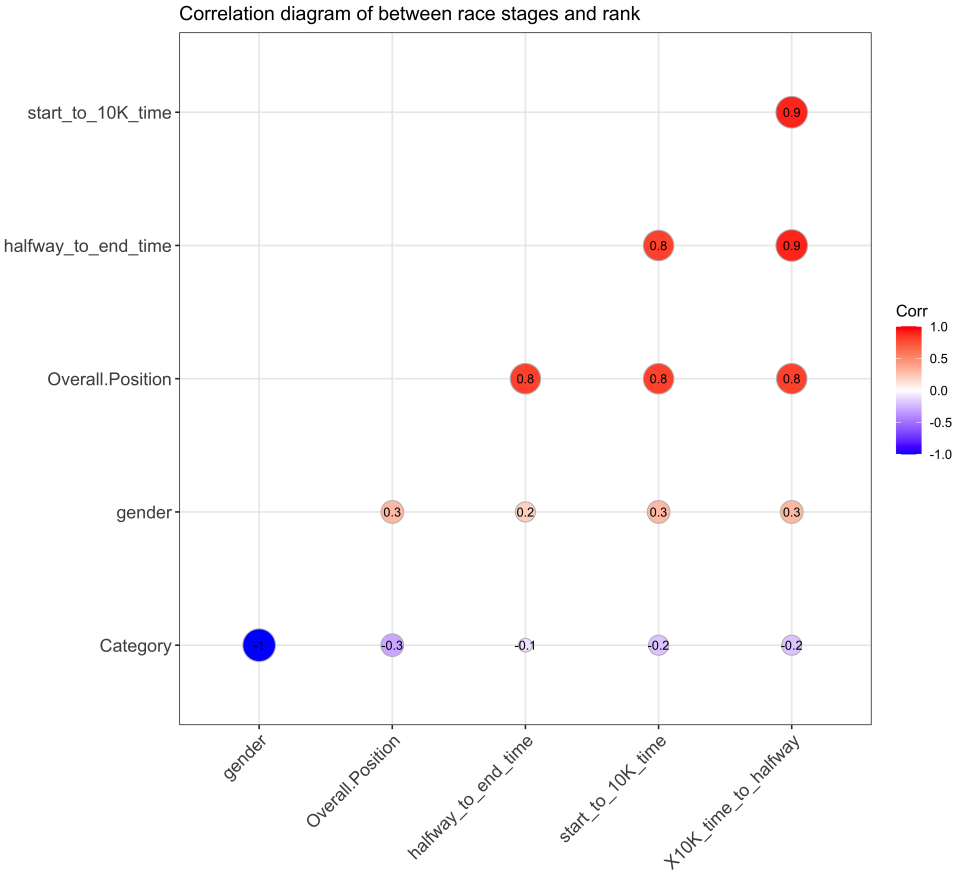


figure below shows type of participants , type can be runner, jogger and walker

based on race finish time , participants can be put into type runner is finish time was below 3hrs , jogger if finish time was between 3hrs and 5hrs and walkers if time was more than 5hrs

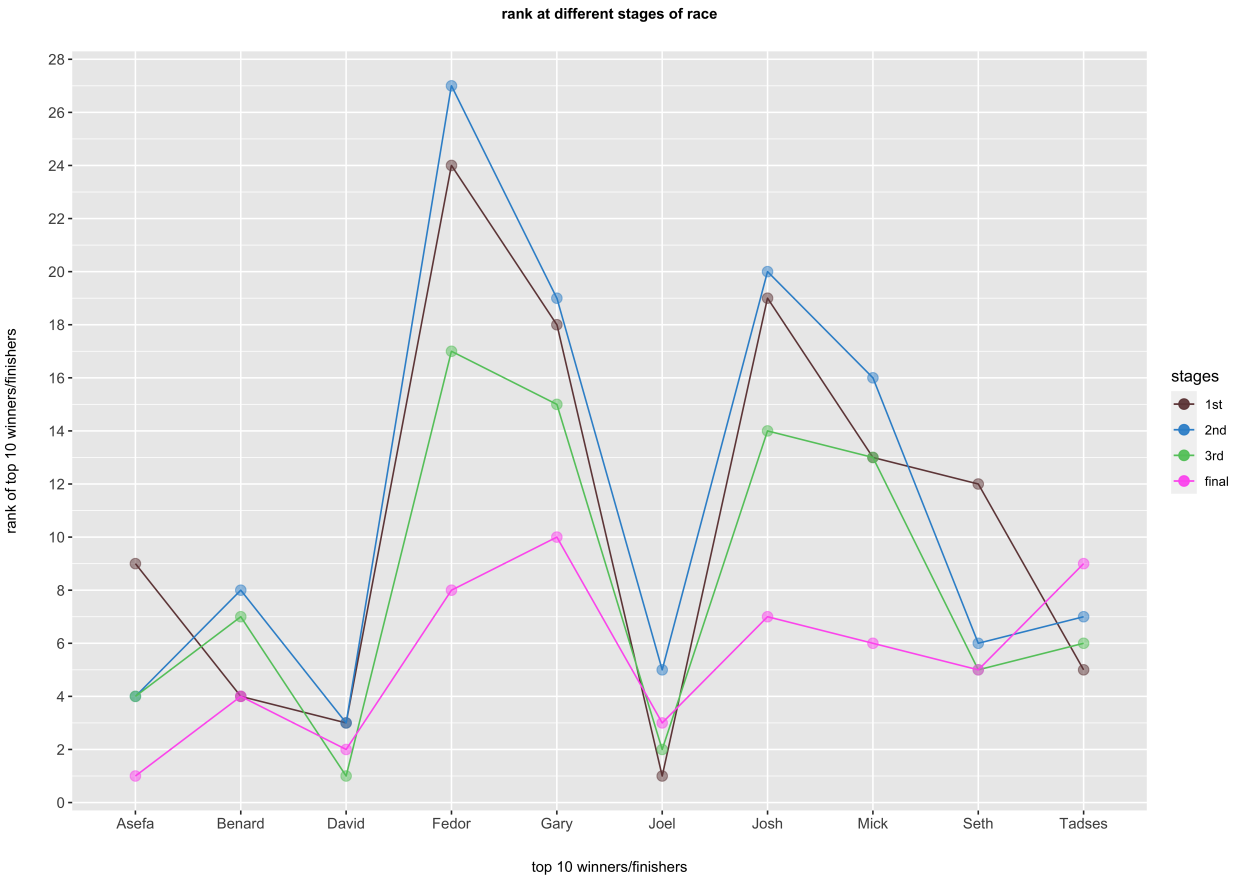


##figure below shows how storngerare variables like time between start to first 10 km , time between 10km to halfway and time bwteen halfway to end correlated with gender and Category *conclusions can be drawn that, fast finishers of at early stages of races are more likely to have better overall position, 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 in the last stage , he has high chances of winning future marathon as his performace is more consistent



Link to code and files