Twin Cities Marathon

Haley Archer-McClellan
hirehaleymcclellan@gmail.com
12/07/2020

Agenda

1. About the Twin Cities Marathon
    a. Why the Twin Cities Marathon? (Read: Race Day Brags)
2. Data Collection Process
    a. Disclaimer
3. Data Processing, uh, Process
4. Findings
    a. Regressions!
    b. Visualizations
        i. Is Marathon Participation Declining?
            1. The Chicago Marathon Effect (maybe)
            2. Sponsorship Effects (Visualized)
        ii. Geography
        iii. Repeat vs One-Time Participants
5. Next Steps

TWIN CITIES IN MOTION

About the Twin Cities Marathon

The Twin Cities Marathon was established in 1982 and has been run every year since then, except for 2020.

The race runs 26.2 miles from Minneapolis to St. Paul, MN

I ran this race in 2018 and wanted to see what factors impact race time based on several years of available race data.

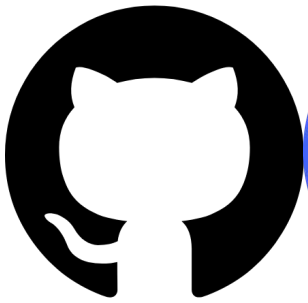
I chose the Twin Cities Marathon specifically because the race data is available online in (essentially) TSV format back to 2001.

Data Collection Process

A year's race results are available through MTEC, a race timing company. For 2015 through 2019, I copied and pasted 500 results at a time into a text document, saved that document as a .tsv once I collected all the results, then opened it in excel and saved as .csv

I found sponsorship data using the Wayback Machine

Data Processing... Process



geopy

I wanted:

1. To anonymize the data so I could host this project on my blog with a clear conscience
2. Distance from the start line (measured as distance from Minneapolis)

Scripted this process:

1. Hash names
2. Use Mapbox to get latitude/longitude
3. Use geopy to get distance to Minneapolis for each lat/long

Regressions

Who runs this race under 3:30? Finding out through regressions:

1. Does state influence race times?
    a. Does the number of times you've run TCM between 2015-2019 influence probability of running the race in under 3:30?
    b. Times Run vs State vs Sex
    c. Times Run as an Integer or Factor
2. Year of Participation

Regressions

- States
  - Positive factors:
    i. Colorado
    ii. DC
    iii. New Mexico
    iv. Washington
    v. Ontario
  - Negative factors:
    i. Minnesota
- Women/Men
  - Women run this race more slowly than men - men and women both run more slowly than NAs (ha, ha)

Regressions

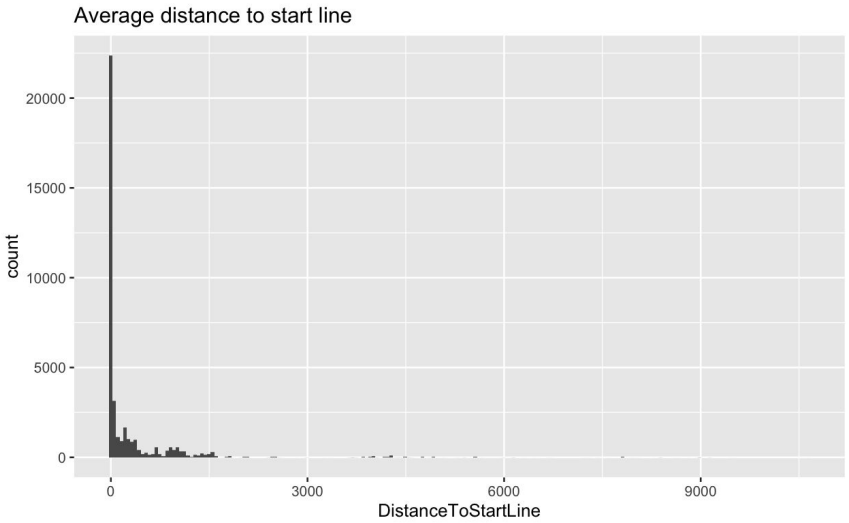
- Times you've run this race:
  - Positively correlated
    i. Overall, positively (statistically significantly) correlated
    ii. 2, 3, 4, and 5 are positively correlated
  - Negatively correlated
    i. 10 is negatively correlated
- Race Year:
  - 2017 was the slowest year - explored weather-related explanations of this

Is Participation Declining?

- We don’t have enough *years* of data to make a meaningful interpretation of this
- Does holding this event the same weekend as the Chicago Marathon cause decreased TCM participation?
  - We don’t have enough years of data to make a meaningful interpretation of this
- Does fewer sponsors imply fewer participants?
  - You guessed it - we don’t have enough years of data to make a meaningful interpretation of this!

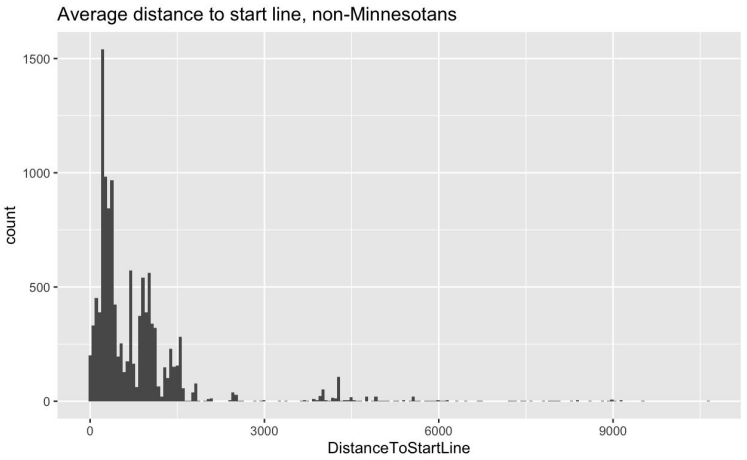
Participant Geography

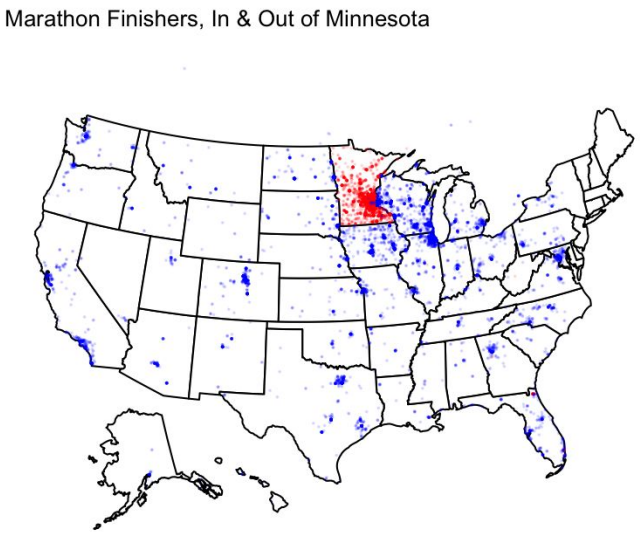
The majority of participants live in Minnesota

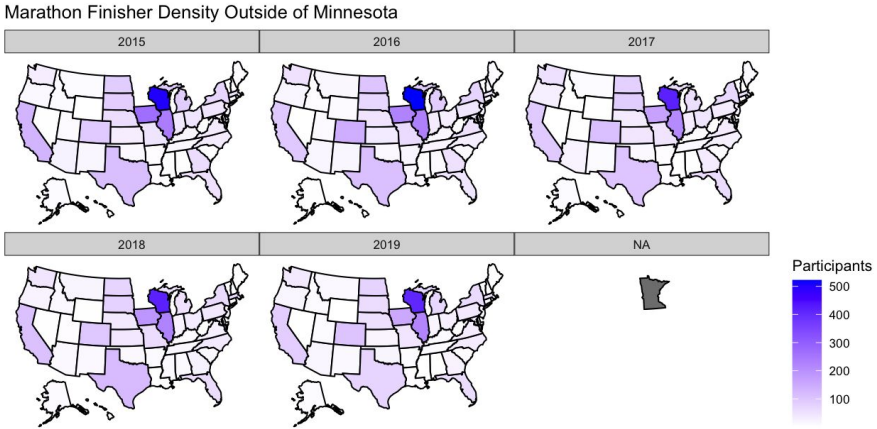


Participant Geography

Outside of the states bordering MN, most participants come from major US cities

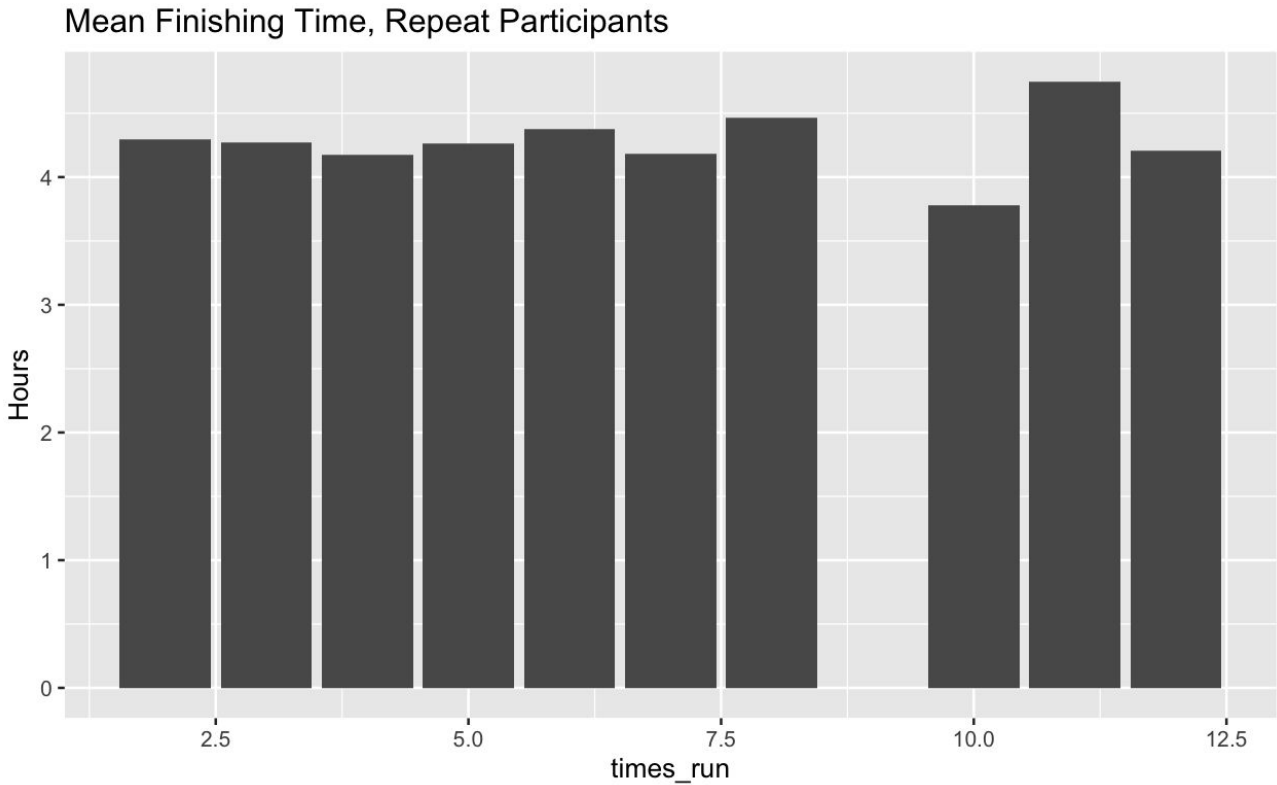






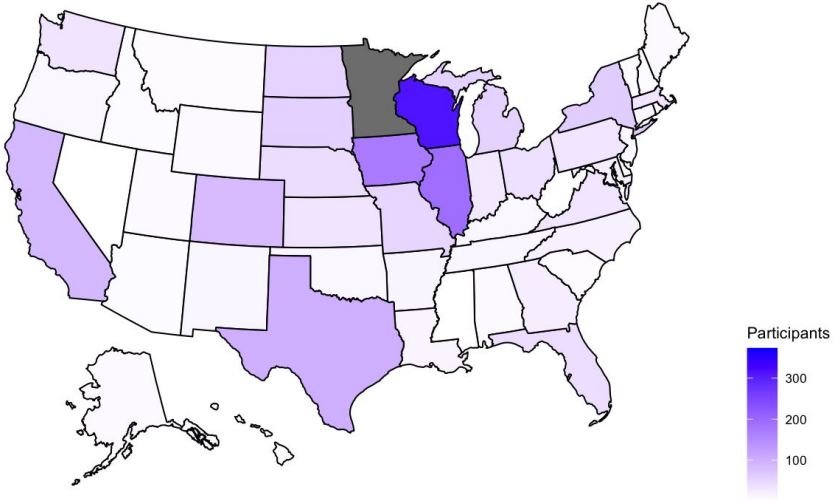
Repeat Participant Finish Times

When doing this analysis, I thought it was odd that someone would have run this event so many times, not thinking about how I only have five years of data - some of these are clearly the same names of different people (ie, John Smith) where there was more than one person by that name running the event in the same year. Oops!

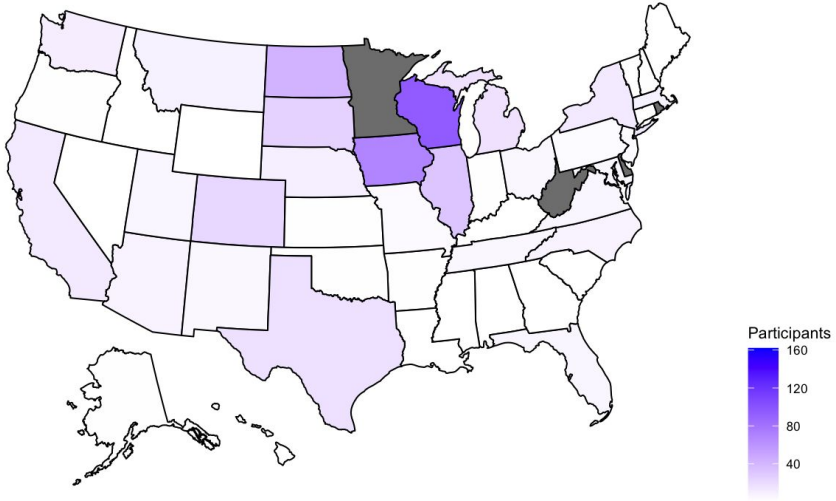


One-Time Participants vs Repeat Participants

One Time Twin Cities Marathon Participants Density, Minnesota Removed, All Years



Repeat Twin Cities Marathon Participants Density, Minnesota Removed, All Years



Future Work

- With more years' worth of data:
  - Study the link between sponsorship and participation in this and subsequent years
  - Study the possible impact of holding the event on the same week as the Chicago Marathon or different years
  - Study the ongoing (seeming!) decline of participation
- With more time to dedicate to my scripts:
  - Improve analysis of one-time vs repeat participants
- More maps!
  - The maps were the most fun part - if I had more time for this project I would build world maps of participants or, with other races' data, maps of other events participants ran

Sources

About the Twin Cities Marathon

Race results

Scripts