

TIME SPENT RUNNING AT AMHERST COLLEGE CROSS COUNTRY PRACTICE

STAT 231: Calendar Query

David Metacarpa

Last updated November 12, 2021

How much time do Amherst cross country runners actually spend running?

I am a member of the cross country team at Amherst College, and while we do spend a lot of time running, we also spend a lot of time at practice doing other things, such as stretching, discussing race strategies, and chatting about whatever comes to mind. Thus, the question arises: just how much time at practice do we actually spend running? More specifically, what percentage of our practice time is spent running? In addition, we have four types of practices: easy days, workout days, long run days, and meet days. At what types of practices do we spend the highest proportion of our time running, and at what types of practices do we spend the most amount of time running? Does this line up with the total amount of time spent at that practice? Finally, how much time do we spend at each type of practice overall, both moving and total?

In order to answer these questions, I collected data on my total time spent at practice by recording the time I arrived at the gym each day and the time I left for my dorm room. I also recorded each of my runs (and other related activities) on Strava, a social media platform designed to help endurance athletes keep track of their activities. Strava automatically uploads each of my runs from my Garmin GPS watch to its website, allowing for me to easily find the time I started and stopped running. Plus, since Strava is primarily used for runners to keep a public journal of their runs, I recorded the type of practice on each of the activities.

This first visualization is a bar chart that compares the total time and the moving time that I spend at each practice, based on the type of practice. Both are average values. The two values are positioned next to each other, with moving time being colored red and total time being colored blue. This chart answers our question of “what types of practices do we spend the most time running at?” Firstly, it is clear that, if I have a meet on a given day, I will spend significantly more time at practice that day than I would if it was an easy day. Despite this, I know that I will only spend a few more minutes running than I would if I had an easy practice or a workout practice, and I will not run as much as on a long run day, even though long run practices take just over half the time a meet day practice does.

```
# Code for bar chart

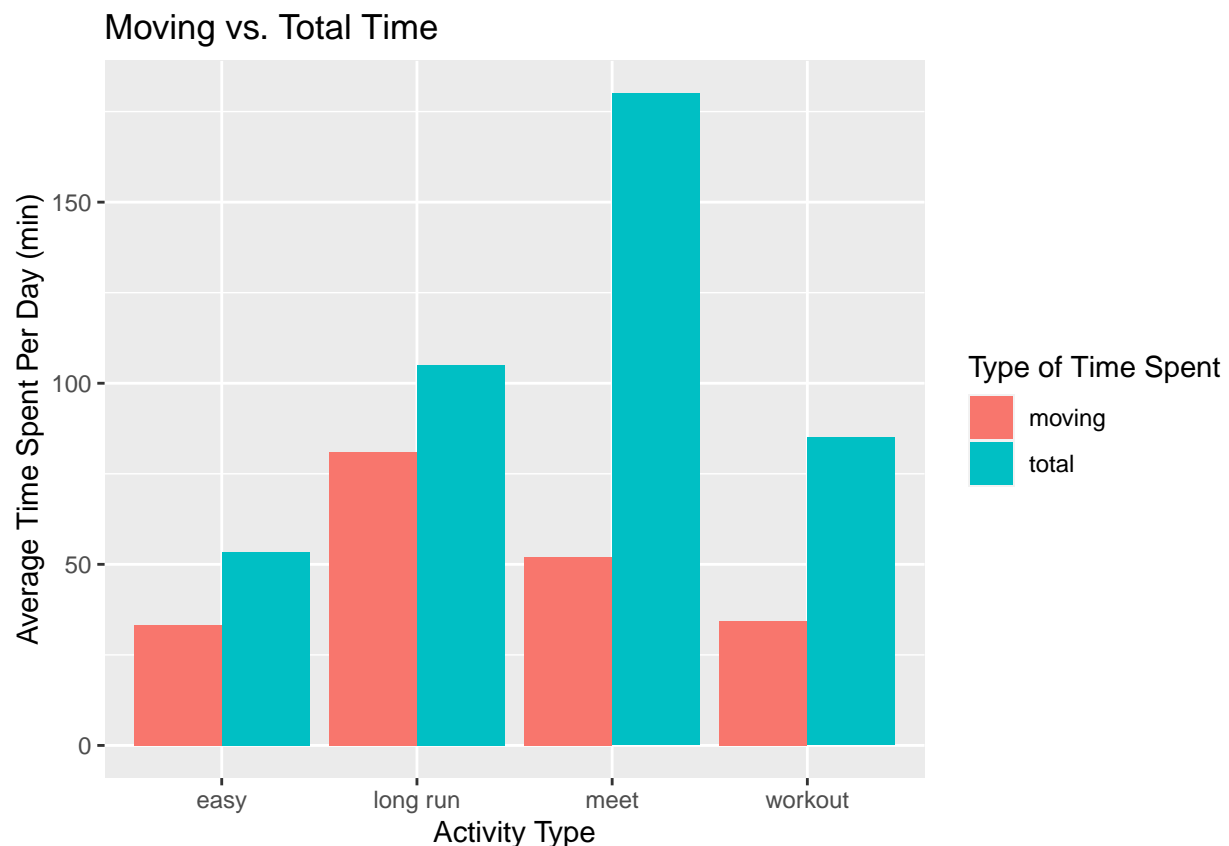
# Data wrangling specific to this bar chart
activities_bar <- activities %>%
  group_by(activity) %>%
  summarize(
    N = n(),
    total_time = sum(duration_min)
  ) %>%
#adding column again for moving vs. total
```

```

mutate(time_type = if_else(activity %in% c("easy moving", "workout moving",
                                           "long run moving", "meet moving"),
                           "moving", "total")) %>%
#adding column again for type of practice
mutate(day_type =
  case_when(
    activity %in% c("easy moving", "easy practice") ~ "easy",
    activity %in% c("workout moving", "workout practice") ~ "workout",
    activity %in% c("long run moving", "long run practice") ~ "long run",
    activity %in% c("meet moving", "meet practice") ~ "meet"
  )
) %>%
#adding column for average time spent at practice
mutate(avg_time = total_time / N)

# Creation of bar chart
ggplot(data = activities_bar, mapping = aes(x = day_type)) +
  # Set position to "dodge" to line the bars up next to each other
  # instead of stacking
  geom_bar(aes(weight = avg_time, fill = time_type), position = "dodge") +
  # Labels
  labs(x = "Activity Type", y = "Average Time Spent Per Day (min)",
       fill = "Type of Time Spent", title = "Moving vs. Total Time")

```



This second visualization is a text plot of the total moving time at each type of practice versus the total time spent at each type of practice. This plot answers our question of “what types of practices do we spend the most time at, both running and total?” It also gives us a visual clue to which types of practices we spend a higher proportion of our time running at, due to the existence of the line segment in the middle of the graph. Firstly, it is clear that we spend more time easy running than any other one type of running, and we spend more time at easy practices than at any other one type of practice. A similar statement can be made for meet days, but for the least amount of time rather than the greatest. Finally, we can conclude that the proportion of time spent running on easy days and long run days is higher than the average, while this proportion is lower than the average for workout days and meet days. This is indicated by the fact that the former two data points are placed above the proportion line segment, while the latter two are placed below the line segment.

```
# Code for text plot

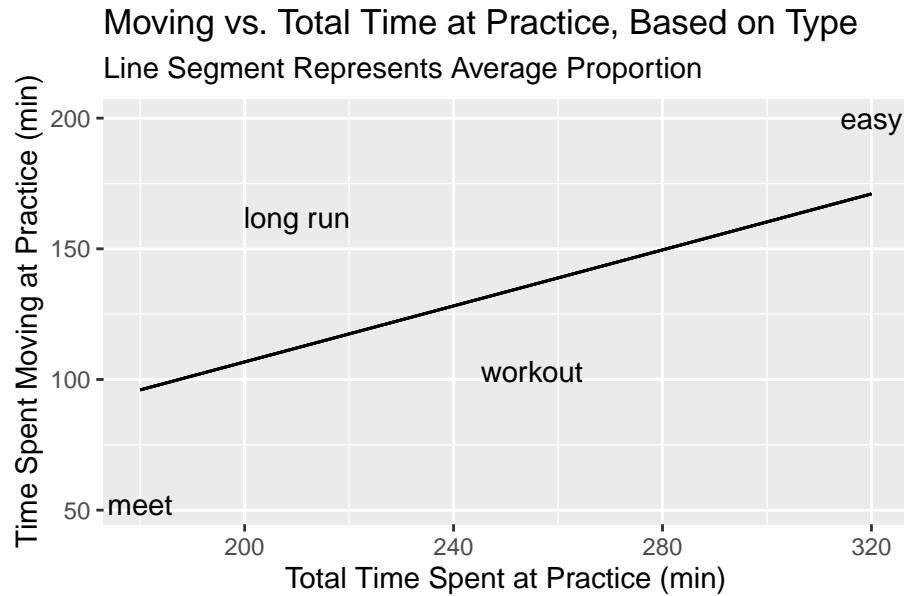
# Data wrangling specific to this text plot
activities_prop <- activities_bar %>%
  group_by(day_type) %>%
  summarize(
    # Creating variables for moving time and total time for each type of
    # practice
    moving_time = first(total_time),
    practice_time = last(total_time)
  )

# Wrangling table again to obtain overall proportion of moving time
# vs. total time
activities_prop_total <- activities_prop %>%
  summarize(
    # Creating sum variables for total time and moving time
    total_moving = sum(moving_time),
    total_practice = sum(practice_time)
  ) %>%
  # Calculating overall proportion and adding it as a variable
  mutate(total_prop = total_moving / total_practice)

# From `activities_prop_total`, we know that the average proportion of time
# spent running at practice is about 0.5357. We will use this information in
# the creation of our visualization, as we will be adding a line segment that
# has the slope of the average proportion.

# Creation of text plot, along with proportion line segment
ggplot(data = activities_prop, mapping = aes(x = practice_time,
                                              y = moving_time)) +

  # Text component
  geom_text(aes(label = day_type)) +
  # Line segment component
  geom_segment(x = 180, y = 96, xend = 320, yend = 171) +
  # Labels
  labs(title = "Moving vs. Total Time at Practice, Based on Type",
        subtitle = "Line Segment Represents Average Proportion",
        x = "Total Time Spent at Practice (min)",
        y = "Time Spent Moving at Practice (min)")
```



The table below shows the amount of time spent running, at practice in total, and the proportion between the two for each day during my data collection period. The entries are sorted by the proportion, in descending order. This helps answer the question “at what types of practices do we spend the highest proportion of our time running?” The table follows a general trend: long run days have the highest proportion, followed by easy days, then workout days, and the lowest proportion is held by meet days. The entries are quite distinctly grouped by day type, as there is no overlap between said groups, save one “swap” between an easy day and a workout day. Therefore, the conclusion can be made that the different types of practice days have almost entirely distinct ranges of the proportion of time spent running.

```
# Code for table
activities_table <- activities %>%
  # We want to use data specific to each individual date
  group_by(date) %>%
  summarize(
    day_type = first(day_type),
    moving_time = first(duration_min),
    total_time = last(duration_min)
  ) %>%
  # Adding a column for the proportion of time spent moving
  mutate(prop = moving_time / total_time) %>%
  # Arranging the entries by descending proportion
  arrange(desc(prop)) %>%
  # Renaming columns for neatness
  rename("type of day" = day_type, "moving time" = moving_time,
    "total time" = total_time, proportion = prop)

# Using kableExtra package for professionalism
kbl(activities_table, booktabs = TRUE, digits = 3) %>%
  # Tidying up heading
  row_spec(0, bold = TRUE, angle = 20, color = "white", background = "black") %>%
  # Creating color spectrum on proportion column
  column_spec(5, color = spec_color(activities_table$proportion[1:12], end = 0.7))
```

date	type of day	moving time	total time	proportion
2021-09-19	long run	83	105	0.790
2021-09-26	long run	79	105	0.752
2021-09-17	easy	32	45	0.711
2021-09-23	easy	52	75	0.693
2021-09-25	easy	31	45	0.689
2021-09-14	easy	40	60	0.667
2021-09-22	easy	22	45	0.489
2021-09-21	workout	35	75	0.467
2021-09-15	easy	23	50	0.460
2021-09-24	workout	38	90	0.422
2021-09-16	workout	30	90	0.333
2021-09-18	meet	52	180	0.289

Reflection

During the data collection process, I ended up hitting a few hurdles along the way, the first of which was finding an accurate way to record the start and end times of my runs on workout days, since I usually time those differently and manually upload them to Strava. This is due to my GPS watch's inability to display lap times during a workout; usually, I record all of my laps as separate activities (with inaccurate start and finish times) or I use a different watch entirely and record the laps manually. Thankfully, Strava was very helpful with this, since all of the laps are displayed under each activity on it anyways. Therefore, all I had to do was change how I was recording the workouts to be the same as for how I would record a regular run and calculate my lap times mentally mid-workout.

The second major hurdle I ended up hitting was a lower spine injury during the second week of data collection. This made it impossible for me to run, but I was able to continue to collect data anyways by simulating all of the runs that the team was doing, both on the stationary bike and by aquajogging. This is the main reason why the term “moving time” appears in this analysis rather than “running time”. Due to these substitute activities being simulations of my actual runs, this analysis is not invalidated by my injury.

These two solutions have positive implications for future data collection and analysis because, unless I injure myself to the point where I cannot do **any** physical activity, I am still able to collect as much data for at least six days a week. Therefore, I would be able to conduct a more thorough and complete analysis in the future. Not only do the solutions to the hurdles I faced help with conducting this analysis, but they also allow for future analyses involving the difference in moving time proportion between injured and non-injured athletes, since injured athletes usually do not simulate what their non-injured counterparts are doing in a given practice.

I believe that I have collected enough of my own data to answer my questions, but collecting more data would further validate the findings of this analysis. The obtainment of this data would not be difficult, as all I would have to do would be to continue to log my runs on Strava and record the start and end times of each of my practices. However, conducting an analysis on the entire team would be **very** difficult. In order to collect data for said analysis, I would either need to record the time each individual runner arrives at and leaves practice as well as all of their runs, or I would need to ask them to record this data themselves. The former would be near impossible with a team of twenty-four (about forty if the women's team is included), and while the latter would certainly be possible, it is not a guarantee that each of the members of the team would have the diligence to record their data in such a fashion, and even if they did, there are bound to be a few members of the team who are uncomfortable giving away such data.