At 9:30 a.m. on August 14, 2016, the women's Olympic marathon kicked off in Rio de Janeiro, Brazil, and 156 runners from 80 countries across the world left the starting line en route to their destination, 42.195 kilometers away. Two hours, 24 minutes, and 4 seconds later, Jemima Sumgong of Kenya would be the first to cross the finish line and take home gold; Sumgong was just 3 and a half minutes slower than her previous personal best time in the marathon.

Approximately 21 minutes later, twin marathoners from Germany—Anna and Lisa Hahner—would cross the finish line together, holding hands and celebrating a personal victory. Although the Hahners would finish 81st and 82nd, respectively, well behind the winners of the marathon, Anna Hahner would describe their joint finish as a “magical moment.”

The media quickly picked up the Hahner story as an image of the beaming twins, finishing hand-in-hand, captured a public audience. While many believed the twins’ near-simultaneous finish was a reflection of Olympic spirit, not everyone agreed with this rosy interpretation. The twins’ happy facial expressions at the finish were portrayed as a bit contrived—smiling like Honigkuchenpferde (cookies in the shape of a horse) was the description offered by one editorialist—and the sports director of the German Athletics Federation, Thomas Kurschilgen, stirred up controversy when he suggested that the Hahners’ photo-finish was no coincidence.

Kurschilgen averred that the twins slowed down deliberately to finish simultaneously and create a spectacle, and he justified...
his charge by noting that the Hahners ran the Rio marathon more than 18 minutes slower than their personal best times before the Olympics. The Hahner twins denied Kurschilgen’s accusations, perhaps not surprisingly.

What happened in the women’s Olympic marathon in Rio, and how might we develop a statistical approach that assesses whether the Hahner twins’ finish in the race was coincidental or intentional?

These two interpretations are clearly at odds. If the former, then the Hahners are to be celebrated and their finish treated as an expression of the spirit behind the Olympic games. If the latter, though, then the twins may have violated this spirit by not trying hard enough. It is perhaps too easy for us to write such a glib sentence—neither of us can fathom being able to complete a marathon anywhere in the vicinity of 2 and a half hours—but we nonetheless want to know what the data from the Olympic marathon tell us.

Among female Olympic marathoners, the Hahner twins were not alone in sharing familial ties. The Rio marathon also featured twins from North Korea, Kim Hye-song and Kim Hye-gyong, who posted identical times and finished 10th and 11th in the race, respectively. However, the Kim finish, unlike the Hahner finish, appears devoid of post-race controversy. Moreover, three triplets from Estonia competed in the Rio marathon, although only two, Lily Luik and Leila Luik, finished it, in 97th and 114th place, respectively. The third Estonia triplet, Liina Luik, recorded what is known as a DNF—“did not finish.” Although our focus here is the Hahner twins, we also touch on the Kim twins and Luik triplets.

Marathon Data and Our Research Design

For each participant who started the women’s Olympic marathon, we know several things: personal best marathon time before the 2016 Olympic games; age; split times from the Rio marathon course at 5 kilometers, 10 kilometers, and so forth; and overall finishing time.

We cannot directly observe the effort that an individual put into the race, and we do not know why some runners have DNF results: Some runners may have injured themselves on the course and accordingly dropped out, and others may have stopped running, uninjured, in anticipation of an unsatisfactory result.

Of 156 marathon starters, 133 completed the race and 23 DNFed at various locations throughout the course. The overall DNF rate was thus 23/156 ≈ 0.15, and the relatively small sample size at our disposal yields a relatively wide 95% confidence interval for this rate, namely (0.098, 0.22).

Kurschilgen’s accusation against the Hahner twins has two components: that these two women ran slowly and that they finished simultaneously. We suspect that Kurschilgen would not have expressed ire at the Hahners had they finished in first and second place in Rio, hand-in-hand and with wide grins. Thus, our investigation of the charges that Kurschilgen offered distinguishes between a slow finish and a simultaneous finish.

Our research design is twofold. First, we present visualizations that describe various features of the 2016 women’s Olympic marathon; among other things, our visualizations feature differences between runners’ Rio times and their prior personal best marathon times. The visualizations suggest that the Hahner twins’ pace in the marathon was slow, albeit not excessively so, but that their simultaneous finish was quite unusual given the twins’ differences in abilities (and similarly for the Kim twins). We then turn to a regression-based simulation of the marathon, and our simulations reinforce what we observed in prior visualizations: that the Hahner twins did not run appreciably slowly, yet finished suspiciously close to each other. We return in our conclusion to Kurschilgen’s claims about the Hahners and offer thoughts about their validity.

Visualizing the Olympic Marathon

One way that we might assess whether a woman marathoner’s Rio finish was unusual—or, say, whether two finishes were jointly unusual—is by comparing a runner’s observed Olympic finishing time on August 14, 2016, with a measure of her underlying marathon talent. For the latter, we use prior personal best marathon times. We believe this is a natural measure of marathon talent, but it is not entirely free of complications. For example, personal best times are potentially confounded by the marathon courses on which they were set; some courses, like the Berlin marathon, are known for relatively fast times. In addition, personal best times may be confounded by conditions, such as weather, that vary across races.

Finally, personal best times may not capture Olympic race day idiosyncrasies that could affect individual runners. For example, a runner could have woken up with a minor cold on the morning of August 14, 2016. With these concerns in mind, we use a runner’s half-marathon split time from the Olympic marathon as a
secondary measure of athletic skill in marathoning.

Figure 1 contains two plots describing how finishing times from the Rio women’s marathon varied as a function of athletes’ personal best times and half-marathon split times. The points in the plots are colored by twin/triplet status, and both plots contain dashed lines representing least squares regression fits. The marks along the horizontal axes in both plots indicate times, either personal bests or half-marathon splits, of runners who earned DNF results.

Considering first the relationship between Olympic finishing and personal best times, Figure 1a’s solid 45-degree line is informative. Given the paucity of points (only five of them) below this line, it follows that the vast majority of Olympic marathoners ran slower in Rio than their prior personal bests.

The Hahner twins were definitely on the slow side, well above the 45-degree line, but a number of runners had even greater differences between their Olympic times and their personal bests. These runners are denoted with squares in Figure 1a, and there are 13 such symbols, highlighting approximately 10% of the racers who completed the Rio marathon. Moreover, the figure shows that two women had personal best times slightly faster than the Hahners and yet finished after the two German women. Although Figure 1a suggests that the Hahner twins were slower than one would have expected given their previous best marathon times, it is not consistent with the accusation that they dramatically slowed down in the Rio marathon.

With respect to our second measure of marathon skill, Figure 1b shows that there was nothing abnormal about the Hahner twins’ overall finishing times, conditional on their half-marathon splits. As one might expect, a runner’s time halfway through the course is a fairly strong predictor of her finishing time. Here we see that the relationship between the Hahner twins’ half-marathon split times and their finishing times is similar to that of other Rio runners. The two points representing the Hahners are in the middle of the distribution of points and therefore imply that the twins do not seem to have slowed down dramatically after they reached the halfway point of the Rio marathon. This is consistent with the aforementioned Figure 1a and inconsistent with accusations made against the Hahner twins; at least the part of the accusation that focused on their overall pace.

Another perspective on the extent to which the Hahners’ overall marathon finishing times were not unusual can be gleaned...
from Figure 2. This figure presents cumulative distributions of the residuals from the regression models displayed in the two panels of Figure 1. In particular, we calculated the residuals from Figures 1a and 1b, Studentized them, and then arranged resulting Studentized residuals from least to most along the horizontal axes in Figures 2a and 2b.

The cumulative residuals depicted in the two panels of Figure 2 are consistent with our interpretation of the Hahner twins’ Olympic finishing times as not particularly remarkable. With personal best time as a measure of marathon talent as in Figure 2a, the two Hahner residuals are in the right tail of the residual distribution; however, their locations are not extreme: One residual is located at the 87th percentile and the other at the 92nd. While these two residuals are in the tail end of the residual distribution, they are not major outliers that would lead us to think that the Hahner twins’ marathon finishing times were extremely slow. Moreover, the residuals for the Kim twins are similarly unremarkable; these two residuals are in the left tail of the residual distribution, indicating that the Kims ran faster than one would have expected.

Finally, if one relies on half-marathon split times as measures of marathon talent, as in Figure 2b, similar conclusions follow. Neither Hahner twin had a finishing time that was particularly unusual given her half-marathon split, and this applies to the Kim twins as well.

If the Hahner twins did not slow down excessively, might they have run somewhat strategically at the end of the Olympic marathon to generate a simultaneous finish? This visualization speaks to this question.

The personal best times of the Hahner twins were 115 seconds apart and their official finishing times were separated by 1 second. Is such a 115 to 1 compression typical among pairs of runners? Have other pairs of Olympic marathoners had a difference between personal best times of 115 seconds apart and, if so, how close were their finishing times?

Of the 133 marathon finishers, there are \( \binom{133}{2} = 8,778 \) pairs of runners. Of these and ignoring the Hahner twins, 10 had exactly a 115-second gap in personal best times. Differences in finishing times of these 10 pairs, in seconds, are: 36, 93, 172, 319, 379, 459, 552, 671, 675, and 739. In other words, of all pairs of runners in the Rio marathon who had a personal best difference that was equivalent to the Hahner twins’ difference, the twins had the greatest compression based on finishing time.
We can generalize this result by looking at all pairs of runners in the marathon. For all 8,778 pairs of finishers, Figure 3a plots differences in finishing times against differences in personal best times, and pairs of twins/triplets are identified by the same color scheme used earlier (recall that only two of the Luik triplets finished the Rio marathon).

Consider first the Hahner twins. The two German women are effectively located on the horizontal axis because their difference in finishing times is 1 second. However, there are many points above the Hahners’ black dot, and this shows that, conditional on an approximate 115-second difference in personal best times, most marathoner pairs did not have close finishing times like the Hahners. Some pairs of runners with around 115-second personal best differences had finishing time differences of 1,000 seconds, i.e., more than 15 minutes. The points in Figure 3a are not independent, but they provide a rough sense of the dispersion in finishing time differences between runners that one might expect, conditional on differences in personal best times.

Thinking about the accusations leveled against the Hahner twins, Figure 3a suggests that Anna and Lisa Hahner did indeed run with an eye on each other. In fact, the same can be said of the Kim twins, who ran seemingly in lockstep throughout the entire Rio marathon. The North Korean twins had a personal best difference of 53 seconds and a finishing time difference of literally zero seconds. Beyond these twins, eight pairs of Rio runners had a 53-second personal best difference and resulting finishing time differences of 9, 51, 228, 340, 352, 571, 662, and 751. As in the Hahner case, the Kim twins compressed their finishing times—meaning that they finished with less time between them than the difference in their personal bests—more than any other pair of runners with similar personal best differences.

Similar conclusions follow from Figure 3b, which plots pair-wise differences between Olympic finishing times and half-marathon split times. Namely, many pairs of runners had similar differences in half-marathon times as the Hahner and Kim twins, but the vast majority of these pairs did not have close finishing times.

Figure 4 describes each Olympic runner’s status at various split times on the marathon course. Each dot in the figure—colored as before—depicts a recorded split and the number of seconds each runner was behind the race leader at the time. There are more dots at earlier splits due to subsequent DNFs.

The Estonian triplet DNF before the half-marathon split is
evident in Figure 4, which also shows that Lisa Hahner was ahead of her sister through 15 kilometers. The figure contains two red dots representing the North Korean Kim twins, but this is not visually apparent because the Kim twins had identical split times during the entire marathon. This explains using the term “lockstep” to refer to the Kim twins’ marathon pace.

**Simulating the Marathon**

Our visualizations shed a fair bit of light on the Hahner twins’ performance in the 2016 women’s Olympic marathon. In the interest of increasing precision, this question arises: If we take into consideration the twins’ similarities in marathon talent and natural variation in marathon finishing times, what is the probability that they would finish the Rio race at roughly the same time and/or sequentially?

To answer this question requires knowing the counterfactual distribution of potential marathon finishing times that would have occurred had the Hahner twins independently (in particular, of each other) and repeatedly run the Rio marathon, holding constant marathon conditions, the abilities of other runners, and so forth. Access to such a distribution would establish the set of potential outcomes that could have occurred on August 14, 2016, and we could in principle use this distribution to determine the likelihood that a simultaneous finish by the Hahners, or at least a near-simultaneous finish, occurred by chance alone.

If these twins rarely finish the marathon together in such a counterfactual world, then one might be skeptical that their observed finish occurred without some degree of coordination.

Unfortunately for us, but fortunately for the race participants, it is not possible to rerun the women’s Olympic marathon to establish a distribution of potential race outcomes for the Hahner twins. However, we can attempt to simulate this distribution by estimating the distribution of every other runner’s finishing time, conditional on marathon talent, and then drawing from this distribution to calculate the likelihood that, for example, Lisa and Anna Hahner finished the Rio race simultaneously.

To estimate the conditional distribution of each Rio final result, we assume that each runner’s marathon time $Y_i$ is distributed normally with a mean that is a function of the runner’s marathon skill, which we assume is a linear combination of her personal best marathon time $X_i$ and her age $Z_i$:

$$Y_i \mid X_i, Z_i \sim N(\beta_0 + \beta_1 X_i + \beta_2 Z_i + \epsilon, \sigma)$$

where $N(\cdot, \cdot)$ denotes a normal distribution. We estimate $\beta_0, \beta_1, \beta_2,$
and \( \sigma \) using ordinary least squares on finishing Rio marathoners. We exclude the Hahner/Kim twins and Luik triplets from the sample so our estimates are not affected by the twin/triplet finishes which, theoretically, could reflect runner coordination.

For a simulated marathon, we draw a runner’s time from an estimated distribution and condition on the runner’s personal best marathon time and age. Once a race is simulated for all runners, twins and triplets included, we record both the time between Anna and Lisa Hahner’s simulated finishes and the difference in their simulated ranks. We then simulate a new race—drawing a new set of finishing times—and record the same quantities. These are the steps in the simulation.

1. Ignoring twins and triplets, estimate a linear model with least squares that predicts a runner’s finishing time \( Y_i \) based on her pre-Olympic personal best time \( X_i \) and her age \( Z_i \).

2. Extract the resulting coefficient vector, estimated covariance matrix, and estimated regression variance from this model.

3. For each simulated race, draw intercept and slope estimates \( \hat{\beta}_0, \hat{\beta}_1, \) and \( \hat{\beta}_2 \), respectively, from a multivariate normal distribution with mean equal to the previously estimated coefficient vector and covariance equal to the previously estimated covariance matrix.

4. For each runner, draw an error \( \hat{\epsilon} \) from a normal distribution with mean zero and a standard deviation equal to the standard deviation of the original regression model’s residuals. This step requires 156 draws from a normal distribution; one draw per marathoner.

5. Predict each runner’s final result by combining the randomly generated beta coefficients and individual error terms, \( \hat{\eta} = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 Z_i + \hat{\epsilon} \).

6. Eliminate each runner from the simulated race with a probability equal to the observed fraction of marathoners who did not finish the marathon.

7. Repeat above steps 10,000 times.

As these steps illustrate, our simulation repeatedly draws random coefficient vectors, and this captures uncertainty in what we know about the relationship between runner talent and age and runner finish times. In addition, for each simulated race, our simulation draws random disturbances for each runner, conditional on the original estimate of regression variance; these disturbances capture variability in runner finishing times, notwithstanding age and underlying marathon talent. Importantly, the disturbances that we draw are independent across runners. Consequently, for each simulated race, the finishing order among runners will vary.

From our simulations, we generate intervals that describe the extent of the variability in marathon finishing times. For example, in 95% of the simulations in which Anna Hahner completed the marathon, her finishing time was between 8,369 seconds and 9,999 seconds. This is consistent with Anna’s observed finishing time in Rio, which was 9,932 seconds. Nonetheless, Anna’s finishing time was on the slower end of this interval. In fact, our simulations estimate that, given Anna’s personal best marathon time and her age, she would be expected to finish the Rio marathon in 9,181 seconds, which is slightly more than 12 minutes faster than her actual time.

This is similar to her sister’s result. Lisa Hahner’s corresponding 95% interval is 8,507 to 10,143 with a mean of 9,319. Her actual finishing time was 9,933 seconds.

The bottom line here is that our simulated marathon finishes are consistent with the Hahner twins’ finishing times in that their finishes were inside the bounds of traditional 95% intervals. And note that the regression model underlying the simulations was estimated without the Hahner twins (and same for the Kim twins and Luik triplets). According to our simulation, then, both Hahner twins did not run appreciably slowly, conditional on personal best times before the Rio Olympics and age.

With an eye on the matter of simultaneous finishing, Figure 5 contains two histograms based on simulated race results. Figure 5a is a histogram that shows the distribution of absolute differences in Hahner twin finishing time where differences are grouped in 30 second bins; counts for the various bins are denoted by the vertical lengths of the bars. Figure 5b is similar, but depicts the distribution of the absolute differences in Hahner twin rankings. Differences are grouped as single units, ranging from no runner between the twins to nearly 120 runners between them.

The histograms in Figure 5 raise questions about the credibility of Anna and Lisa Hahner’s story and, in particular, suggest that a simultaneous finish in the Rio marathon...
would be very rare if Anna and Lisa had run independently. For example, in fewer than 300 of 10,000 simulated races did Anna and Lisa Hahner finish within 30 seconds of one another, and in fewer than 600 did the Hahner twins finish within a minute of each other. The histogram area associated with this latter result is depicted in red in Figure 5a.

Moreover, the Hahner twins finished in consecutive rank in fewer than 200 of 10,000 simulated races; the red zone in Figure 5b presents this visually. The close finish that was observed in Rio, where Anna and Lisa crossed the finish line one after the other, would have been highly unlikely if the two German twins had raced independently of each other.

Parallel to our prior analyses, we repeated our simulations using half-marathon splits as the predicting variable in our simulations; results are in Figure 6. While this
use of marathon splits reduces the variation in the predicted outcome of each runner and therefore reduces the expected distance between Anna and Lisa Hahner, even with half-marathon split as a measure of ability, it is still quite rare for the twins to finish simultaneously or consecutively.

The way that we handled DNFs in our simulations is notable. As our description indicates, we assumed that DNF probabilities are the same for all runners and that the likelihood of a DNF is not a function of a runner's anticipated marathon finishing time. The rug marks in Figure 1a suggest that runners with better personal best times may be more likely to DNF than other runners, all things equal. We suspect that this occurs because some better runners may expend excessive energy trying to achieve a good result in the marathon and in so doing, injure or exhaust themselves; lesser runners, in contrast, may be content to finish respectably.

Regardless of the validity of this conjecture, Figure 1a shows that the Hahner twins are representative of the sort of runners who DNFed in Rio. Since our simulated Hahner statistics are conditioned on both Hahner twins finishing, it follows that they are conservative.

The fact that both women finished the marathon was notable in and of itself and, by discounting the possibility of a Hahner DNF, we are giving the benefit of the doubt to the Hahner twins.

**Conclusion**

Anna and Lisa Hahner’s near-simultaneous finish in the 2016 women’s Olympic marathon in Rio elicited a controversy in that the German twins were accused of deliberately slowing down and finishing next to each other to generate media attention. They denied this, not entirely surprisingly.

Of one perspective on the marathon that is based on visualizations and simple calculations, and a second that draws on a simulation, both have the same implications. In a global sense, the Hahner twins did not slow down appreciably during the Rio marathon. Their times were not fast, but they were within reason for runners of the Hahners’ abilities and age. Locally, though, we find that the Hahners’ finish probably was, in fact, contrived. Their finish—consecutive, with 1 second between the two women—was a rather low-probability event. Compared to their differences in talent, that is, the Hahner twins difference in finishing times was unusually compressed. This is evidence that their finishing had elements of intentionality—perhaps at the last minute, but intentionality nonetheless.

Our goal is not to speak to whether the Hahner twins should or should not have enjoyed what some might call an artificial moment of Olympic glory. Neither was in contention for a podium finish in Rio, and compared to the doping allegations that presently surround endurance sports in general, what the Hahners appear to have done seems relatively tame. Still, it might have behooved them to have been a bit more open about their end-of-race tactics. To this end, it is clear how a simple data analysis can shed light on claims about racing results.

Should the 2020 Summer Olympics again feature marathoning twins and triplets, we look forward to a comparative analysis using the techniques illustrated here.

### Further Reading


---

**About the Authors**

David Cottrell is a lecturer in the Department of Government at Dartmouth College and was a postdoctoral research fellow in Dartmouth’s Program in Quantitative Social Science when he completed work on this article. He teaches courses on the application of data analysis and statistics in the social sciences and is currently involved in research that leverages computational approaches to study the effect of gerrymandering.

Michael C. Herron is William Clinton Story Remsen 1943 Professor of Government and chair, Program in Quantitative Social Science, at Dartmouth College and was a visiting scholar at the Hertie School of Governance, Berlin, Germany, when he completed work on this article. He has taught courses in applied statistics at Dartmouth since 2004. At present, he is researching relative age effects in professional football and the extent to which Americans vote infrequently, thus exposing themselves to the risk of being removed from registered voter pools.