

## Me vs The Elites



This picture tells the story of me being a guineapig at my local Sports Science Lab a few years ago. It is at this moment I acquired an interest in data, sort of.

I responded to an advert on [social media](#) and subjected myself to the [Sports Scientists](#) under the leadership of [Dr Tim Noakes](#) who wanted to research what separated an elite from a sub-elite athlete. After many experiments and gruelling sessions, they reported what I already knew, I am a 'good club rider'. Which is to say that I am somewhere between a couch potato and an elite bicycle racer. The researchers used [watts/kg](#) after an initial short Functional

Threshold Power test ([FTP](#)) to determine who they assumed to be elite and sub-elite. Their research also revealed what we all already suspected; that performance depends on the body and the mind. The 'mind' being that part of you that informs the decisions you make. The sports scientists elaborated that elite athletes are motivated to defeat everybody else (outside of themselves) whereas the sub-elites tend to want to beat only themselves (a sort of internal battle).

I have decided to obtain some data (by populating an excel document). The timekeepers, [RaceTec](#) keeps valuable data of all your races which they have timed. There are several timekeepers, but I focused my attention only on the local events timed by RaceTec. I removed the mountain bike events and focused only on road racing results. With this [data](#) I proceeded to compare myself to the overall winner. I assumed that the overall winner was an elite. I did not take into account my age vs the age of the winner. I did not compare myself to the winner in my category or the age grading. I ignored things like mechanical breakdowns, punctures and crashes in the race in which I was involved in. (There's no data on that, neither can I remember). RaceTec uses an algorithm to place you into a category based on your speed in previous races as well as the course. Hilly courses score a higher ranking than flat ones. Higher your seeding and therefore the faster the group generally go as fast as the pack you're in, a sort of mass

To clean the data, I changed the values in 'event\_name' comes on board the name of the event changes. The group used integers instead of characters for 'my\_start\_group' on speed and distance and not on the finishing time.

| Power Profile (Men) (Watts/kg) |       |       |      |      |
|--------------------------------|-------|-------|------|------|
|                                | 5 s   | 1 min | PPO  | FTP  |
| World class<br>(Pro Tour)      | 24.04 | 11.50 | 7.60 | 6.40 |
|                                | 23.69 | 11.35 | 7.47 | 6.29 |
|                                | 23.35 | 11.21 | 7.34 | 6.17 |
|                                | 23.00 | 11.06 | 7.21 | 6.06 |
|                                | 22.66 | 10.92 | 7.07 | 5.95 |
| Exceptional<br>(International) | 22.31 | 10.77 | 6.94 | 5.84 |
|                                | 21.97 | 10.62 | 6.81 | 5.72 |
|                                | 21.62 | 10.48 | 6.68 | 5.61 |
|                                | 21.28 | 10.33 | 6.55 | 5.50 |
|                                | 20.93 | 10.19 | 6.42 | 5.38 |
| Excellent<br>(National Level)  | 20.59 | 10.04 | 6.29 | 5.27 |
|                                | 20.24 | 9.89  | 6.16 | 5.16 |
|                                | 19.90 | 9.75  | 6.02 | 5.04 |
|                                | 19.55 | 9.60  | 5.89 | 4.93 |
|                                | 19.20 | 9.46  | 5.76 | 4.82 |
| Very good<br>(Elite)           | 18.86 | 9.31  | 5.63 | 4.71 |
|                                | 18.51 | 9.16  | 5.50 | 4.59 |
|                                | 18.17 | 9.02  | 5.37 | 4.48 |
|                                | 17.82 | 8.87  | 5.24 | 4.37 |
|                                | 17.48 | 8.73  | 5.10 | 4.25 |
| Good<br>(Top League Racer)     | 17.13 | 8.58  | 4.97 | 4.14 |
|                                | 16.79 | 8.44  | 4.84 | 4.03 |
|                                | 16.44 | 8.29  | 4.71 | 3.91 |
|                                | 16.10 | 8.14  | 4.58 | 3.80 |
|                                | 15.75 | 8.00  | 4.45 | 3.69 |
| Moderate<br>(Competitive)      | 15.40 | 7.85  | 4.32 | 3.58 |
|                                | 15.06 | 7.71  | 4.18 | 3.46 |
|                                | 14.71 | 7.56  | 4.05 | 3.35 |
|                                | 14.37 | 7.41  | 3.92 | 3.24 |
|                                | 14.02 | 7.27  | 3.79 | 3.12 |
| Fair<br>(Well Trained)         | 13.68 | 7.12  | 3.66 | 3.01 |
|                                | 13.33 | 6.98  | 3.53 | 2.90 |
|                                | 12.99 | 6.83  | 3.40 | 2.78 |
|                                | 12.64 | 6.68  | 3.27 | 2.67 |
|                                | 12.30 | 6.54  | 3.13 | 2.56 |
| Untrained                      | 11.95 | 6.39  | 3.00 | 2.45 |
|                                | 11.61 | 6.25  | 2.87 | 2.33 |
|                                | 11.26 | 6.10  | 2.74 | 2.22 |
|                                | 10.91 | 5.95  | 2.61 | 2.11 |
|                                | 10.57 | 5.81  | 2.48 | 1.99 |

course. Hilly courses score a higher ranking than flat ones. It is interesting to note that the faster you are the higher your seeding and therefore the faster the group you will have for company. In cycling you shall generally go as fast as the pack you're in, a sort of mass hypnosis.

To clean the data, I changed the values in 'event\_name' to a more common name. Whenever a new sponsor comes on board the name of the event changes. The groupby() function gave me some problems. I have used integers instead of characters for 'my\_start\_group' so that an A group = 1, B = 2, C = 3 etc. I focused on speed and distance and not on the finishing time.

## The Data in General

From the describe() function we learn a few interesting ideas. My average speed was 31.8km/h while the elites clocked in at 40.1km/h. My standard deviation is 3.7 while the elites are at 2.3. This shows that they are more consistent while my results are less predictable. I think most athletes would prefer consistency. My slowest speed was

|       | distance_km | my_avg_speed_kph | my_start_group | winners_avg_speed_kph |
|-------|-------------|------------------|----------------|-----------------------|
| count | 144.000000  | 144.000000       | 144.000000     | 144.000000            |
| mean  | 101.326389  | 31.880069        | 7.819444       | 40.110833             |
| std   | 28.609214   | 3.760671         | 6.391267       | 2.349913              |
| min   | 40.000000   | 22.020000        | 1.000000       | 33.250000             |
| 25%   | 85.000000   | 28.752500        | 3.000000       | 38.662500             |
| 50%   | 97.000000   | 32.625000        | 6.000000       | 39.975000             |
| 75%   | 109.000000  | 34.830000        | 11.000000      | 41.617500             |
| max   | 202.000000  | 39.990000        | 26.000000      | 46.570000             |

22.0km/h while the elite's crawling speed was 33.2km/h. Max for me was just under the elite's average of 40km/h while the elite's max was 46.5km/h. It seems like I am slower by a whopping +/-10km/h overall. The difference between my slowest and my fastest is about 18km/h. The elite's difference is about 13km/h. My range is wider compared to the elites. In cycling you can decrease your weight, increase your power or both. ['Training load'](#) is a common term used in sport's science. It is a vague idea but generally you have to put

yourself under moderate pressure over a particular time. You decide if you want to increase strength or speed or fitness. The beauty of being an amateur is that you are in charge of you. But let us assume that you decide to work on any particular aspect and improve by a reasonable 2% per annum.

By examining my performance on speed from 2005 to 2021 I went from a maximum average speed of 29.51km/h to a maximum of 37.99km/h. This is an increase of approximately 26% divided by 15 years equals almost 2% improvement. For me to get to the elite's average speed will take me:

|            | distance_km | my_avg_speed_kph | my_start_group | winners_avg_speed_kph |
|------------|-------------|------------------|----------------|-----------------------|
| event_date |             |                  |                |                       |
| 2005-04-12 | 71.0        | 28.15            | 26.0           | 35.48                 |
| 2005-10-22 | 44.0        | 28.78            | 26.0           | 38.61                 |
| 2005-10-30 | 40.0        | 29.51            | 26.0           | 38.37                 |
| 2005-11-12 | 72.0        | 24.88            | 22.0           | 33.25                 |
| 2006-01-28 | 80.0        | 24.24            | 26.0           | 41.78                 |
| ...        | ...         | ...              | ...            | ...                   |
| 2020-02-16 | 116.0       | 37.18            | 2.0            | 45.13                 |
| 2020-02-22 | 109.0       | 28.94            | 4.0            | 42.36                 |
| 2020-08-03 | 109.0       | 36.12            | 4.0            | 43.58                 |
| 2021-05-23 | 90.0        | 37.99            | 1.0            | 45.44                 |
| 2021-10-10 | 109.0       | 37.07            | 1.0            | 41.61                 |

$$\text{Number of years} = \frac{\text{winner's avg time} - \text{my avg time}}{\text{my avg time}} \times 100\% \div 2\% = 13 \text{ years}$$

13 years is a terribly long time. I think I'll just enjoy the game without extra pressure because I am also getting older. Even though [training supposedly reverses aging](#). I believe [with age it becomes more difficult for the brain to remain motivated to do strenuous training](#). An interesting side note is that most people do not realise the difference between [exercise and training](#), [health and fitness](#), [strength, power and speed](#).

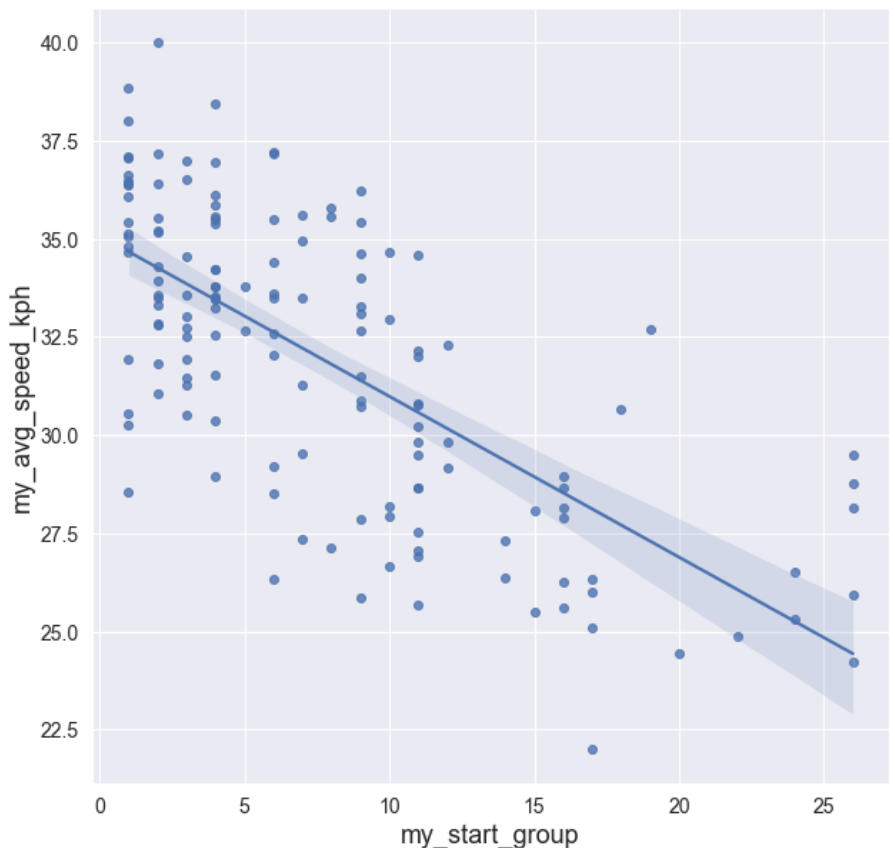
## Heat Map

The heat map shows correlation between features. As previously mentioned, there is a 'mass hypnosis' effect phenomenon in cycling. Looking at the heat map we can see a correlation between my\_speed and my\_start\_group. -0.7 is a strong indicator that the closer you are to the beginning of the alphabet the faster you will go. Recall that I used A = 1 and therefore the correlation is a negative integer. The further back towards the other end of the alphabet, the slower your speed. It is difficult to see the cause-and-effect relationship because you have to get fit and strong and get faster on your own and then join a faster group which will propel you even faster. Please also understand the [drafting effect](#) in cycling. (Note that the elites all start in the same group at the same time. The value for their start group will always be the same. I also did not focus on any particular elite racer, but rather the time of the winner. The

winners change all the time.)

## Regression Plot

The regression line confirms what the heat map suggested, you go faster if you're in a faster group. Using this regression plot you can predict the average speed based on the start group. Recall that A = 1, B = 2, etc. As also mentioned earlier I did not account for mechanical breakdowns, punctures and crashes. If I could recall which races had unfortunate breakdowns or crashes I could remove the outliers and get a more accurate result. However, I think the data does tell a holistic story, accurate enough.





Looking at the graph of my speed vs the winners speed it is obvious that the elites are more consistent. One would assume that the speed would be slower for longer races, but the data shows that the distance doesn't really affect the average speeds.