

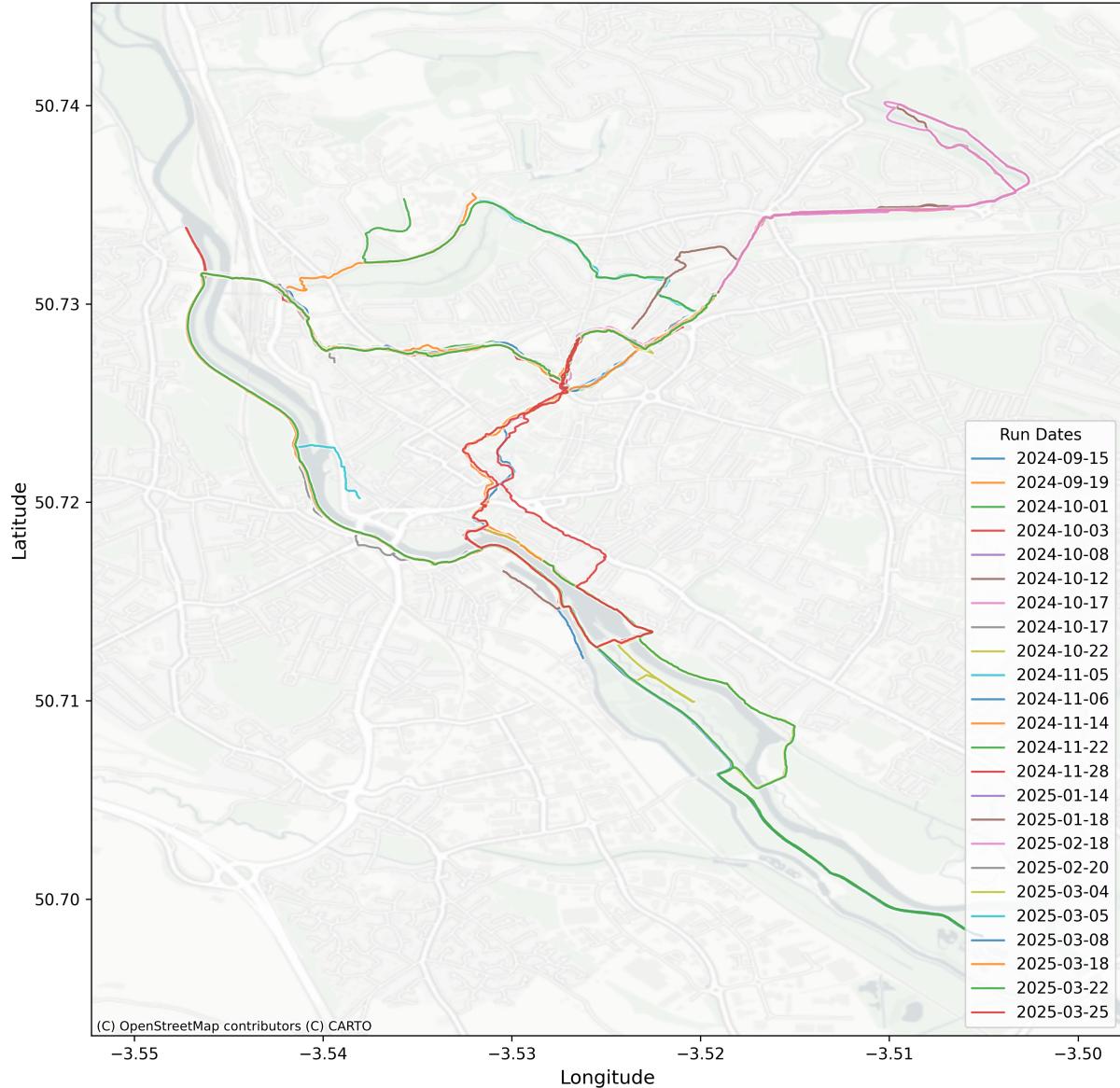
'Run Forrest Run!': An Overkill Analysis of Personal Strava Data

Have you ever looked at the price of Strava Premium and gone 'I could do this sh*t for free'; Well, look no further. At the eye-watering cost of £55 a year, the only reasonable option is to download the entirety of your own personal data from the Strava API and process it yourself - whilst taking some creative liberties. Not only that, but I faced the confusion of my friend, Issy, when I sent her a list of instructions to obtain her data too..

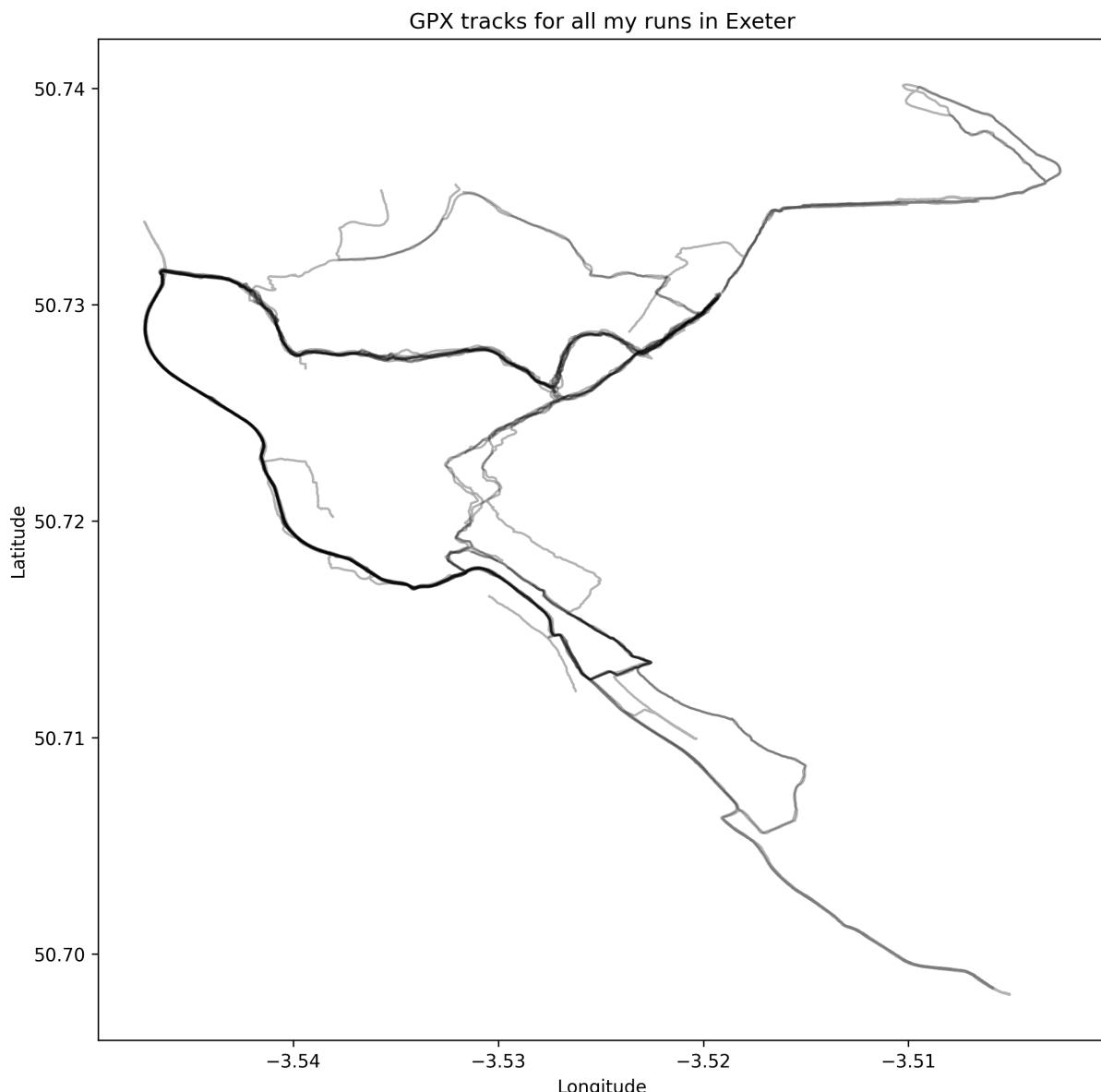
Routes

First up.. where are my generic running routes in Exeter? Strava can't do this one! These two graphs visualise the same set of geo-spatial data, however it is aestheticised in two ways to provide some extra analysis.

GPX Tracks For All My Runs in Exeter with Terrain Basemap



First off, the coloured graph separates out all my runs, presenting a more memorable visualisation of where I end up on my travels. For example, I can pick out the bright green and tedious 11k I did in March along the river. Furthermore, this map is useful for presenting good run routes in Exeter - such as routes down to the Quay, Marsh Barton, or Stoke Hill. It shows that I have done many runs, often favoring the same routes, and perhaps it evidences my address slightly too well.



The latter grayscale graph is a tinkering with the line transparency, revealing the patterns and repetitions I show in route of running. From this, it can be understood that I am pretty predictable - I often go to the quay and along the river as shown by the boldness of the black lines in these parts. However, it seems I stray from this pattern too. Combined with the previous terrain map of Exeter, this is an apt show of all of my personal runs in exeter.

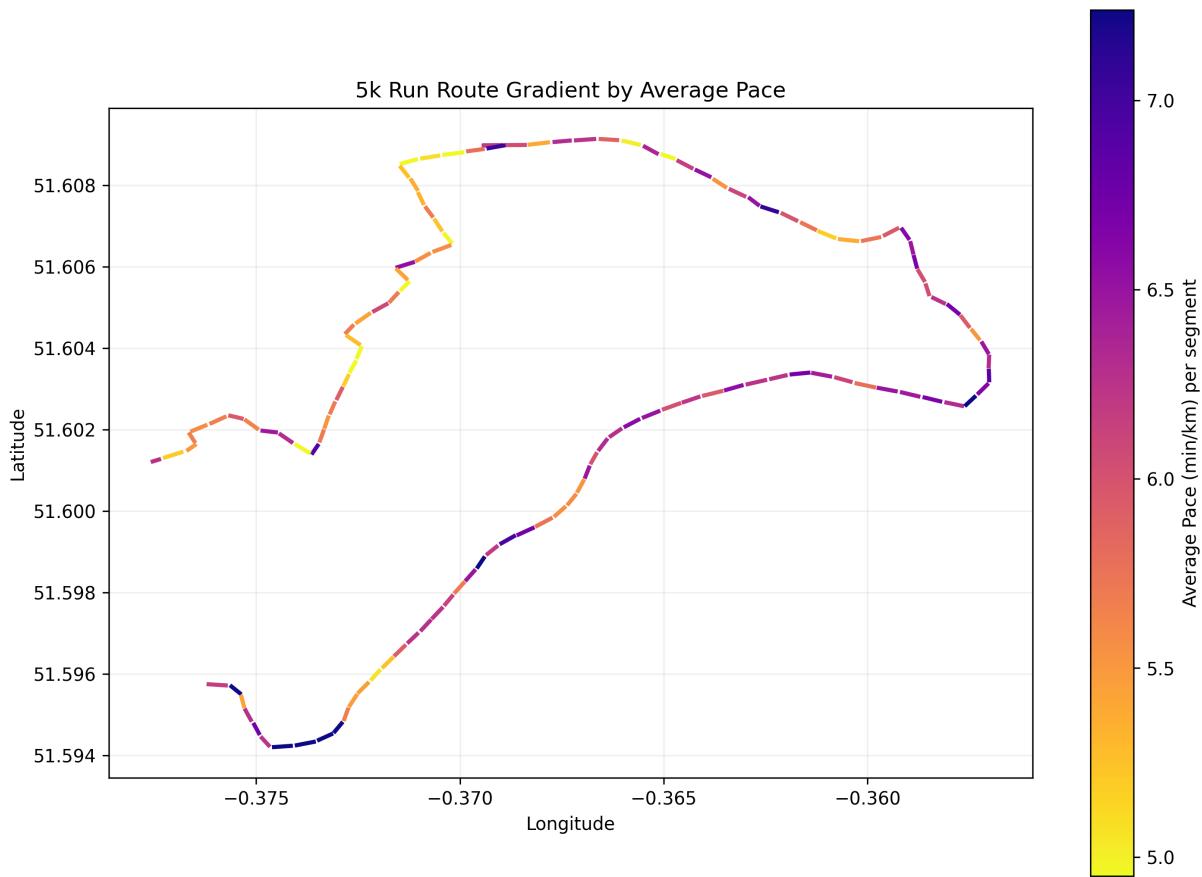
| Where my runs often end:



Exeter Quay, on 'Morning Run, 2024-11-14

Speed

Routes have been covered but what about speed and pacing? I selected a 5k run from my strava to analyse my pace. The run started at the lesser longitude/latitudes.

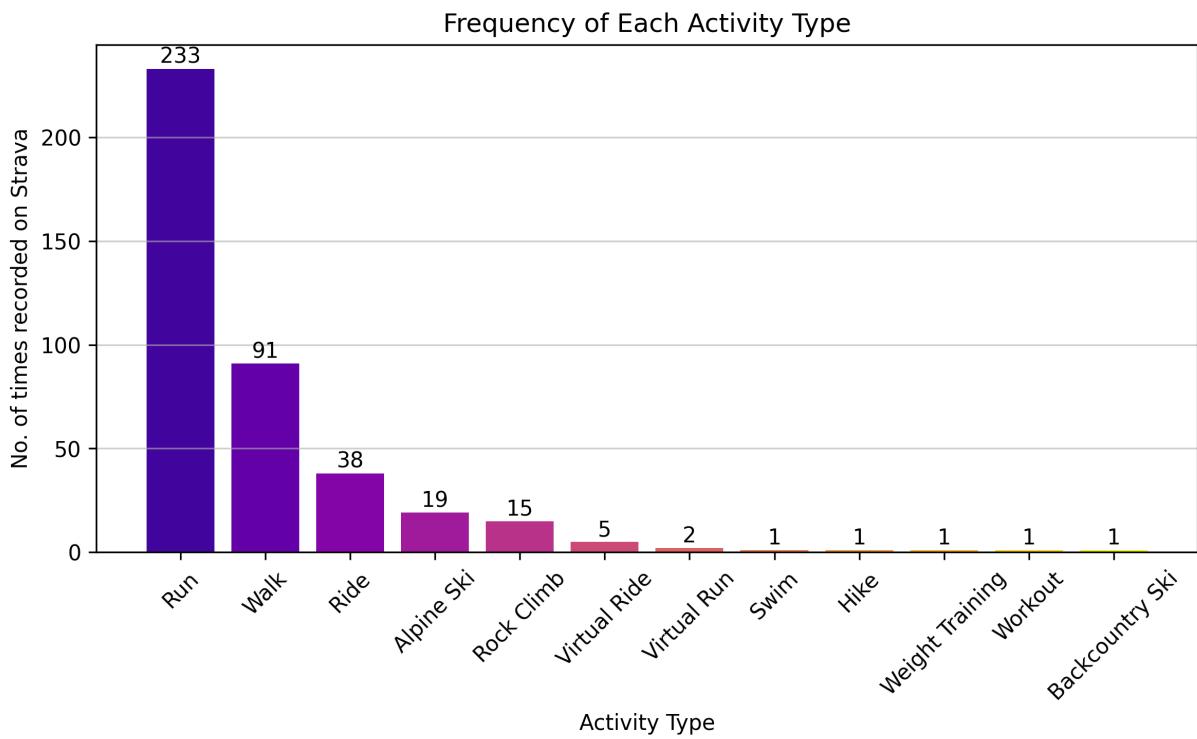


5.00km run, 'a bit wet', 2024-08-24, ran in 28m 35s and overall pace of 5:43/km

This graph, with sections of 40m showing an average pace gradient, shows that I might have poorly planned my pace in this run. It presents that I spent the first half running relatively slowly, at a pace largely above 6 min/km. However, in the second half I spend up relatively significantly, maintaining speeds of 5.5min/km and below. Whilst perhaps considered inconsistent running to maximise overall average speed, the faster ending could indicate a downwards gradient of elevation, or rather, some personal necessity to finish strong. Adding a terrain base map to this graph might better serve analysis. The latter graph is a zoom in of runs below 7km, in order to better analyse the many more data points under this distance.

Introducing Issy

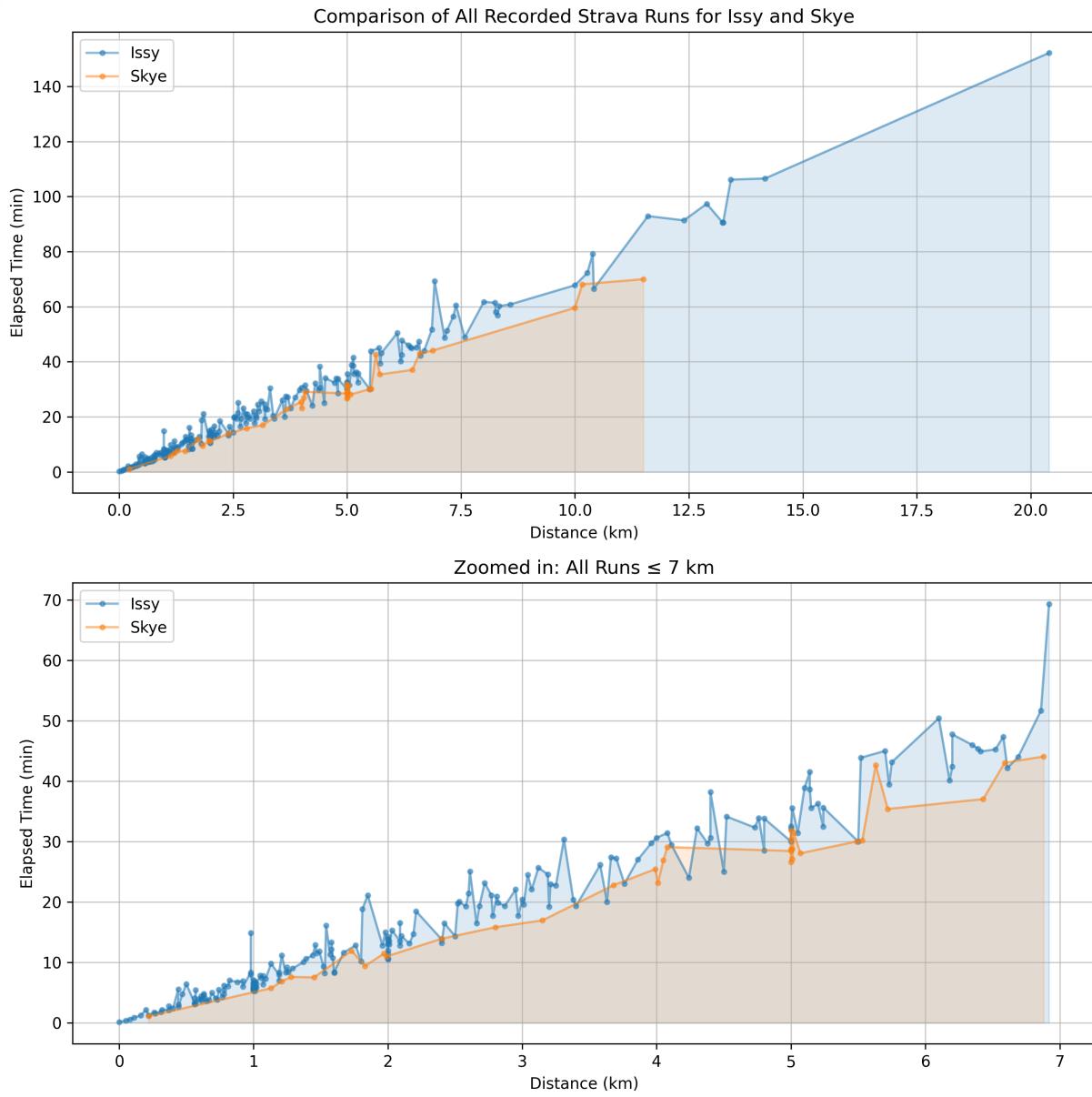
With over 400 recorded activities on Strava, including 233 runs, Issy, also in the same gender age group as me, is an ideal candidate for data comparison. Her breakdown of activities are expressed in the following plot.



This bar chart shows that, alongside, running, she also enjoys walking, cycling, skiing, and rock climbing. However, I will be focusing on running only.

The following graph depicts all of our individual recorded runs on Strava with the time and distance associated with each point. This means that the lower the data points the faster

the run.



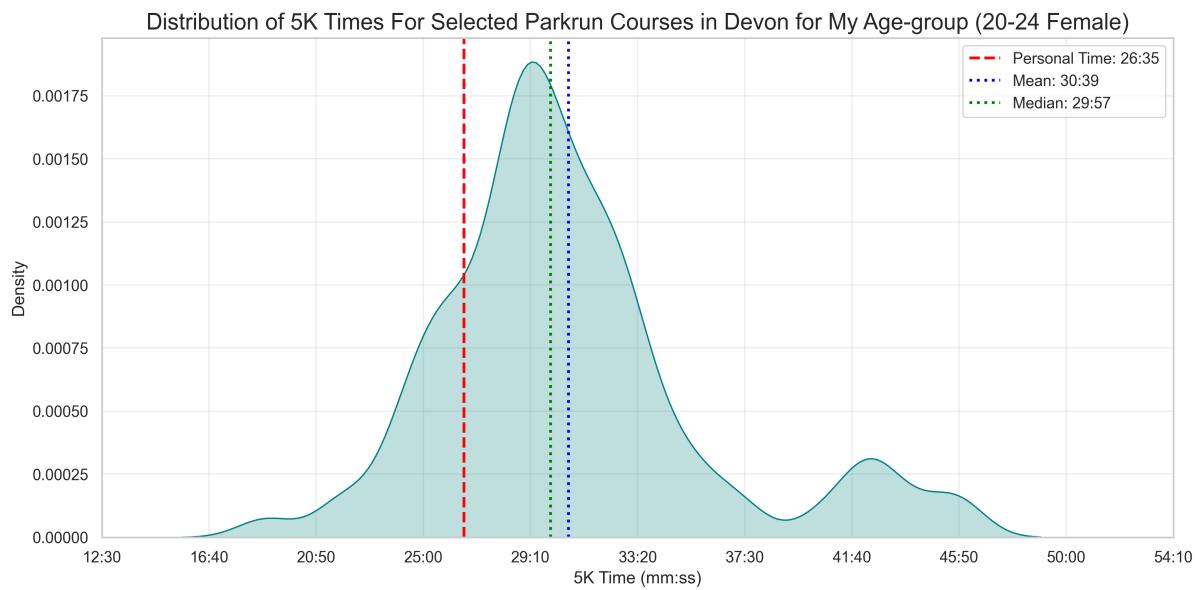
Firstly, the graph shows that Issy has many more recorded activities than me, as shown by the visibly higher count of blue points, particularly clustered between 0km and 5km. Additionally, she has completed runs at a higher distance than me, where the longest run I have completed is around 12km, whilst her longest run is above 20km.

We both display a similar pattern, of a gently slowing pace as distance increases, however, there is volatility among data points, showing we are not too consistent. This might reflect that we purposely don't put maximum effort into every run.

Furthermore, my runs are observedly slightly faster than Issy's, however, she has many more recorded activities, where I only have 36. This makes any speed variations less comparable. Other factors, such as our choice of regular run route or training methods, might also erode the solidity of this conclusion.

Parkruns: a Comparison to Other 5k Runners in Devon

Parkruns are community powered local 5k runs held every Saturday across the country by volunteers, where complete times are digitised and posted on Parkrun's website. I selected 9 locations of parkruns across Devon/Cornwall (Exeter, Exmouth, Cranbrook, Teignmouth, Greendale, Penryn, Plymouth, Mount Edgecombe, Plymvalley) and scraped for results of females in the age 20-24 category.



This KDE plot offers some valuable insights into the variation of times achieved by women in this age group. Firstly, a high kurtosis is observed, with a gentle skewedness to the right, where many participants achieved close 5k times. This is reinforced by a higher mean than median. This skewedness is indicative of times clustering towards the lower end, with a smaller number of outliers to the higher end. These outliers are characterised by the second peak in the data, with times between 39 and 50 minutes. This could be explained as a bimodal distribution, with some portion of participants perhaps attending parkruns for the first time and not returning, lending to the smaller peak, whilst the normal peak might be characterised by regular parkrun attendees. Further, research leads me to understand that average 5k times for this age group across the country is between 36-38 minutes. The average times achieved by parkrun goers is much lower, which implies parkrun attendees are faster than the average runner for this demographic group, at least in Devon.

How does this lend to my personal analysis? This plot would indicate that I display a faster 5k time than the average girl in my age group in Devon.

Conclusions

Whilst I cannot match Strava's data insights quite yet, this project shows the potential of the data downloadable from Strava, offering visualisations not necessarily offered by the premium membership. In terms of results, it is fair to say my running routes can be quite predictable and I favour some routes over others. Furthermore, in comparing my pace to others in my demographic group, evidence suggest my personal results are relatively fast.