

Running Speed Distribution Write-Up

Our dataset consists of results from 992 maximal graded exercise tests (GETs) conducted by researchers at the University of Málaga from 2008 to 2018. We started with two datasets `subject_info.csv` and `test_measures.csv`. The `subject_info` dataset has information about participant characteristics, including age, sex, height, and weight, while the `test_measures` dataset contains cardiovascular and respiratory measurements. We joined these two datasets on 'ID_test' and filtered to include only the first exercise test of each participant to maintain consistency.

Our goal is to explore the question: *How do your age, height, weight, and gender stack up against other runners?* We focused on this question as we wanted to understand how we compare to other runners and what insights we can gain from broader trends. By analyzing these factors, we aim to provide meaningful context allowing individuals to benchmark their own statistics against a relevant dataset. By doing this, we can offer a more curated and accurate comparison for evaluating personal maximum speeds, helping runners gain deeper insights into their performance relative to their peers.

We started by brainstorming how the overall structure of the visualization should look, considering how to best present the data at the base level. Since our goal was to provide an overarching view of the dataset, we explored various graph types, including histograms, line plots, and scatter plots. Ultimately, we decided on a histogram as it offers an intuitive distribution of our data, making comparisons clearer and more effective. We then considered different interactive features to enhance the user experience. Initially, we explored options such as a highlight or zoom functionality that would allow users to view specific data points in greater detail. However, we ultimately decided that implementing a direct filtering system, which allows users to adjust the histogram based on categories within the four selected factors (age, height, weight, and gender), would provide a more meaningful way to explore the data and answer our overarching question. Filtering enables the user to get as specific as they want, ensuring the most personalized and relevant results. We also incorporated a tooltip that appears when the user hovers over a bin, offering more specific details.

For the development process, we thought it would be best to meet in person to complete each aspect of the project, as it would be the easiest and most efficient way to collaborate. We practiced pair programming by working together and taking turns to contribute to the visualization code. We spent around 3-4 hours creating the application. The most difficult part was figuring out how to bring our concept to life and how to incorporate all the visual elements we had in mind.