

## Context

This is a historical dataset on the modern Olympic Games, including all the Games from Athens 1896 to Rio 2016. I scraped this data from [www.sports-reference.com](http://www.sports-reference.com) in May 2018. The R code I used to [scrape](#) and [wrangle](#) the data is on GitHub. I recommend checking [my kernel](#) before starting your own analysis.

Note that the Winter and Summer Games were held in the same year up until 1992. After that, they staggered them such that Winter Games occur on a four-year cycle starting with 1994, then Summer in 1996, then Winter in 1998, and so on. A common mistake people make when analyzing this data is to assume that the Summer and Winter Games have always been staggered.

## Content

The files `athlete_events_part_1.csv` and `athlete_events_part_2.csv` contain 271116 rows and 15 columns. Each row corresponds to an individual athlete competing in an individual Olympic event (athlete-events). The columns are:

1. ID - Unique number for each athlete
2. Name - Athlete's name
3. Sex - M or F
4. Age - Integer
5. Height - In centimeters
6. Weight - In kilograms
7. Team - Team name
8. NOC - National Olympic Committee 3-letter code
9. Games - Year and season
10. Year - Integer
11. Season - Summer or Winter
12. City - Host city
13. Sport - Sport
14. Event - Event
15. Medal - Gold, Silver, Bronze, or NA

## Acknowledgements

The Olympic data on [www.sports-reference.com](http://www.sports-reference.com) is the result of an incredible amount of research by a group of Olympic history enthusiasts and self-proclaimed 'statisticians'. Check out their [blog](#) for more information.

## Inspiration

This dataset provides an opportunity to ask questions about how the Olympics have evolved over time, including questions about the participation and performance of women, different nations, and different sports and events.

Source:

<https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results/>  
home