In his book *Outliers*, Malcolm Gladwell argues that the timing of a hockey player's birth can give them a lifelong advantage. Gladwell claims that 40% of elite hockey players are born between January and March, due to the cutoff date for Canadian age-class hockey programs. Children born earlier in the year are usually larger and more physically developed than their younger competitors, which results in them being identified as better athletes and receiving more coaching.

In this analysis, we will be investigating if this phenomenon also appears in athletes competing in this year’s Olympics. Specifically, we will compare the birth month of Olympic athletes competing in individual events with those competing in team events. For team events, we will investigate soccer, volleyball, and basketball. For individual events, we will investigate artistic gymnastics, swimming, and track and field.

**Obtaining the Data** (see footnote)

To carry out this analysis, we will be using two websites, (1) the Olympic roster to get the athlete names for each sport, and (2) Wikipedia to get birthdays for each athlete.

* First, you will need to grab the list of athletes from the Olympic roster. Fortunately, the website has separate pages for each sport, which follow consistent formatting (https://olympics.com/en/paris-2024/athletes/<SPORTNAME>).
  + <https://olympics.com/en/paris-2024/athletes/artistic-gymnastics>
  + <https://olympics.com/en/paris-2024/athletes/football>
  + <https://olympics.com/en/paris-2024/athletes/swimming>
  + <https://olympics.com/en/paris-2024/athletes/volleyball>
  + <https://olympics.com/en/paris-2024/athletes/basketball>
  + <https://olympics.com/en/paris-2024/athletes/athletics>
* Now that you have the names of the athletes you are interested in, you need to obtain their birth month from Wikipedia. Fortunately, Wikipedia also follows consistent formatting for individual profiles ([https://en.wikipedia.org/wiki/<FIRSTNAME>\_<LASTNAME](https://en.wikipedia.org/wiki/%3cFIRSTNAME%3e_%3cLASTNAME)>)**\***.
  + The born field on each athlete’s card tells their birth month—this is what you want to keep.

**\*** Note that we are okay with removing any athlete whose Wikipedia page cannot be accessed by inserting their name from the Olympic roster into the generic URL format.

**Analyzing the Data**

Now that you have data on the month each athlete was born, we will compare these birth months within and between individual and team sports.

* Create a table of birth months for the individual and team sport groups (aggregating over events).
  + *Within* the individual and team sport groups, does it appear that athletes born at a certain time of the year have an advantage?
  + *Between* the individual and team sport groups, are there differences when athletes were born?