

# How have the Olympic games' athletes changed over time?

## Introduction

**Business Context.** You work for a company that specializes in analyzing data for a variety of clients in the sports industry. Some questions that you frequently encounter include determining if a new player is promising enough to invest money in their development, which teams are the most likely to win certain matches, what events will be the most attractive to advertisers, etc.

**Business Problem.** As part of one of your projects, you have been asked to perform an exploratory data analysis of historical data to **detect patterns in the provenance, physical profile, and other characteristics of the athletes who compete in the Olympic games**. The conclusions of your analysis will help the rest of the team prepare a report for a new client who helps sports gear manufacturers find advertising opportunities.

**Analytical Context.** You have scraped a dataset from the Internet, which contains data for all the Olympic games from Nagano 1998 to Rio 2016. It comprises data for 46,349 individual athletes and has 13 columns for each one of them. There are 69,847 rows instead of 46,533 rows in the `olympics_data` worksheet because some athletes have won multiple medals:

- **ID:** A unique number assigned to each athlete
- **Name:** The athlete's name
- **Sex:** The athlete's sex
- **Age:** The athlete's age at the moment of the games
- **Height:** The athlete's height in centimeters
- **Weight:** The athlete's weight in kilograms
- **Team:** The athlete's team (country)
- **Year:** The year
- **Season:** The season
- **City:** The host city
- **Sport:** The sport the athlete competed in
- **Medal:** The medal that the athlete won, if any (can be Gold, Silver, Bronze, or NA)
- **Won medal?:** 1 if the athlete won a medal, 0 otherwise

The dataset can be downloaded from this link  
(/extended.basic\_eda\_excel\_lite\_fellow/files/data/olympics\_fellow\_rubric.xlsm).

**Instructions:** For this case, you will need to submit the Excel file with your work on it.

- Please write all your answers (formulas and text response) in the `ANSWER` worksheet.
- Each question will have a designated area where you can write your answer.
- Further instructions will be given when you reach those questions.

**Note:** Some individuals have competed in more than one Olympics game, which means that there are duplicate athletes. You don't need to take that into account when solving the exercises.

## Height, weight, and age

### Exercise 1

#### 1.1

Calculate the average height, weight, and age of athletes in Rio 2016 across all sports.

**Instructions:**

- Write your formulas in cell `AB8` , `AB10` , and `AB12` in the `ANSWER` tab.

**Answer.**

#### 1.2

Repeat Exercise 1.1 but for Sydney 2000.

Have the averages changed noticeably?

**Instructions:**

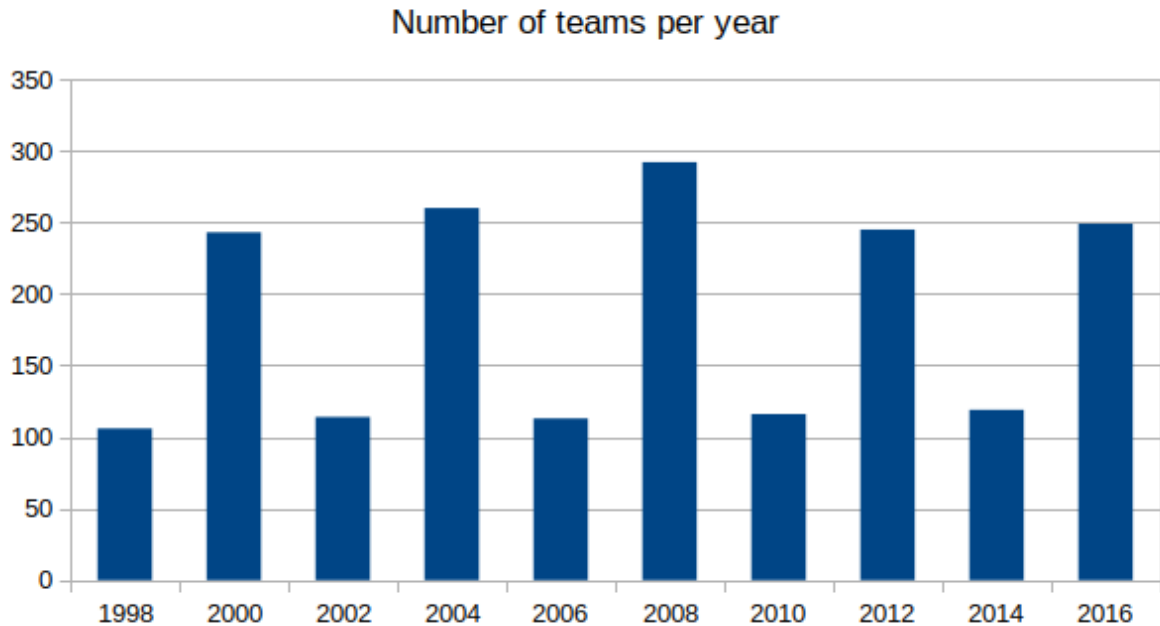
- Write your formulas in cells `AB51` , `AB53` , and `AB55` in the `ANSWER` tab.
- Write your written answer in cell `AR49` in the `ANSWER` tab.

**Answer.**

## Geographic representation

### Exercise 2

This is a chart of the number of countries that participated in the games from 1998 to 2016.  
What can you conclude from it?



**Hint:** Keep in mind that Summer and Winter games are not held in the same year. In the Winter games, the number of teams is typically lower than in the Summer games.

**Instructions:**

- Write your written analysis in cell X94 in the ANSWER tab.
- Respond to the question with at least two sound arguments that are supported by the graph.
- Use at least two evidence-based examples to support arguments. Your paragraph should be logical, easy to understand, and generally free of spelling errors.

**Answer.**

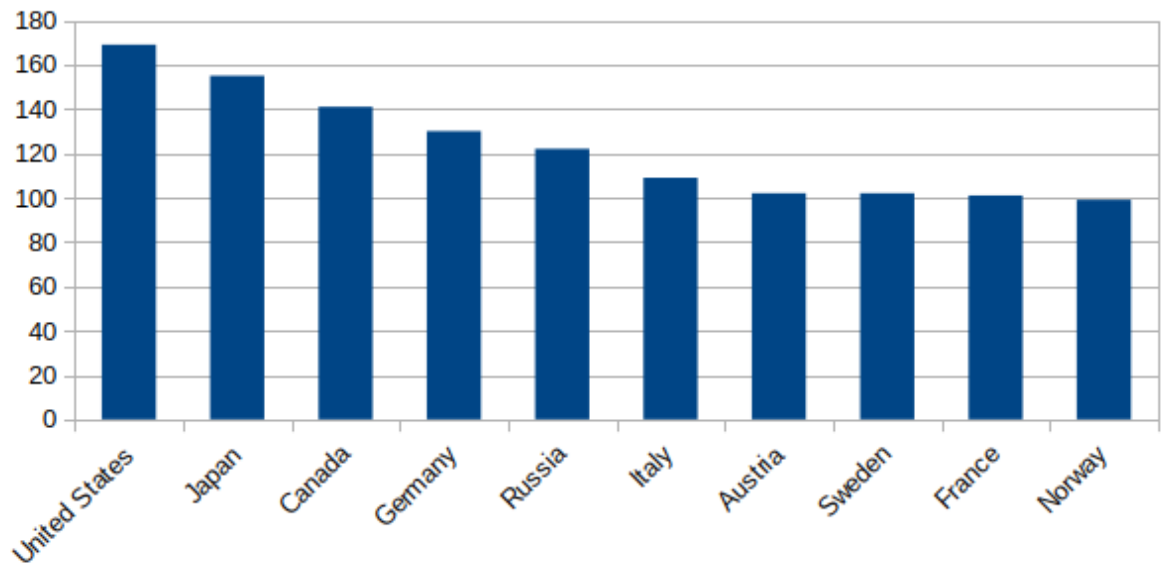
### Exercise 3

Below are bar charts for the top 10 countries by number of athletes sent for all the games between 1998 and 2016. Examine the bar charts and look for similarities and differences in the data over the years and the seasons (Winter vs. Summer). What trends do you notice?

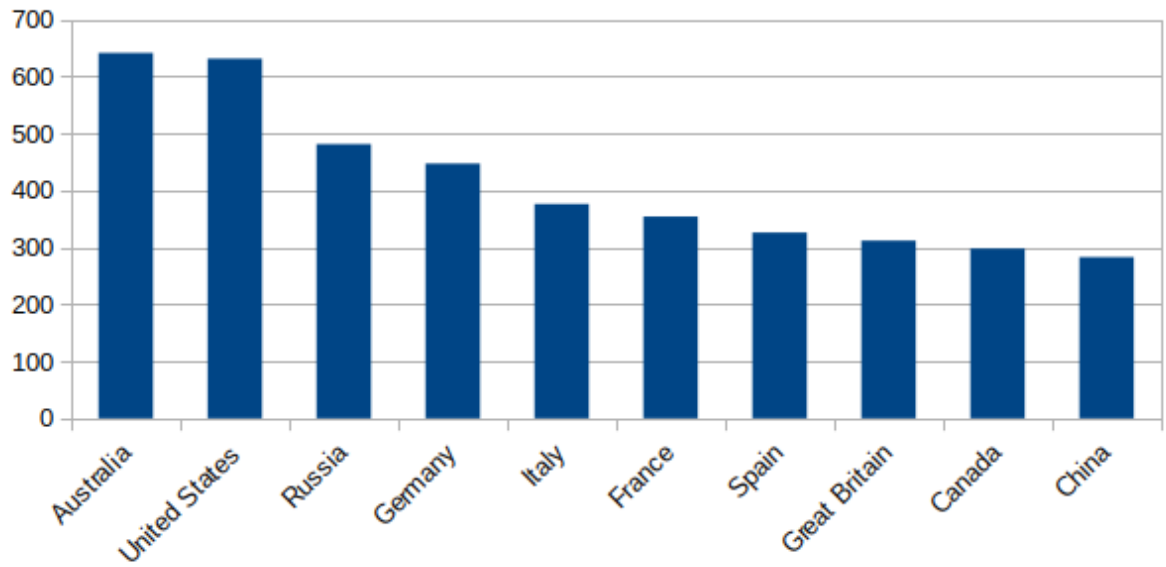
**Instructions:**

- Write your written analysis in cell X126 in the ANSWER tab.
- Respond in at least three sentences. At least two claims should be made and backed with evidence from the graph to support them. Responses should be clear, with minimal spelling and grammatical errors.

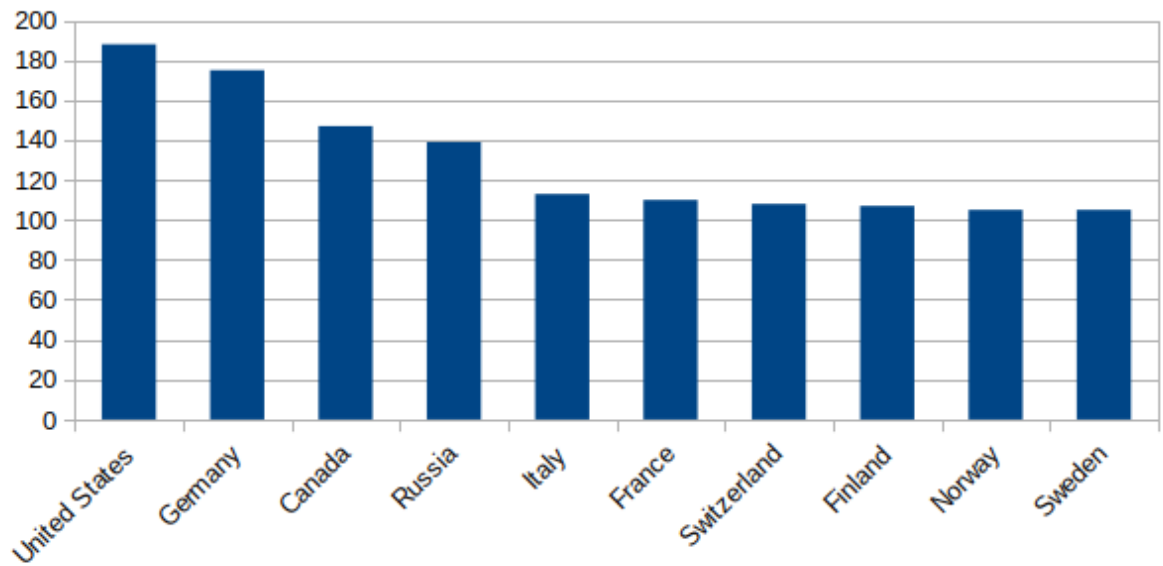
Top 10 1998 (Winter, Japan)



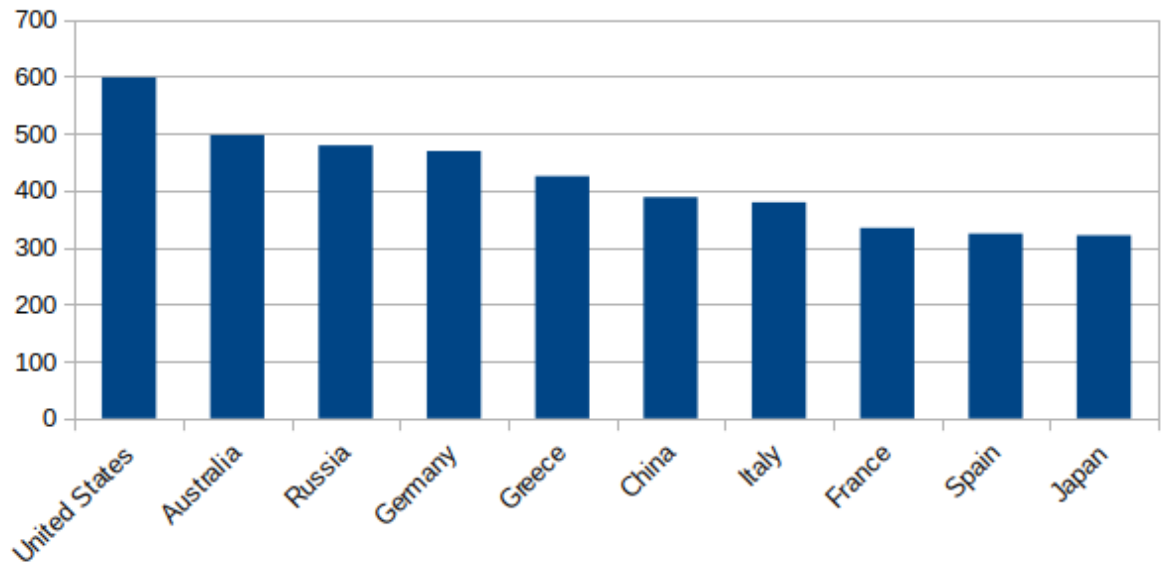
Top 10 2000 (Summer, Australia)



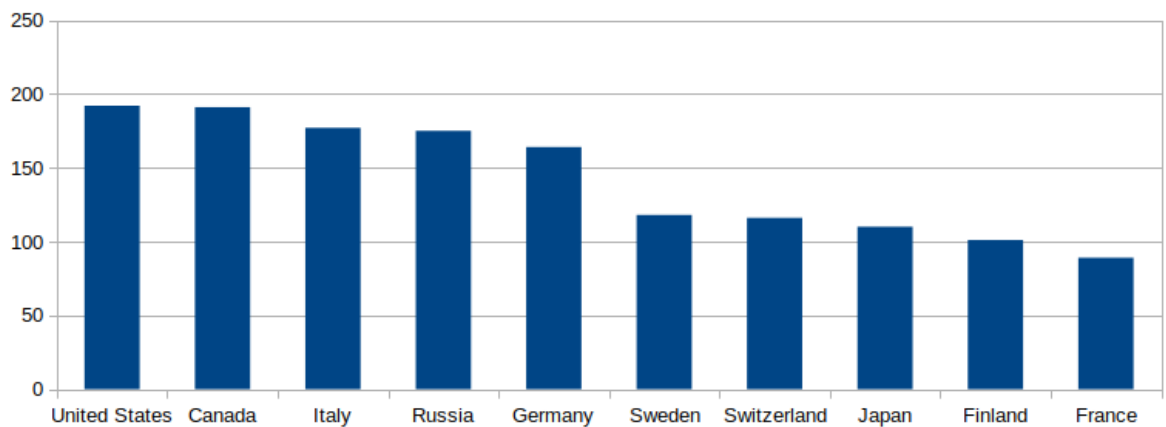
Top 10 2002 (Winter, USA)



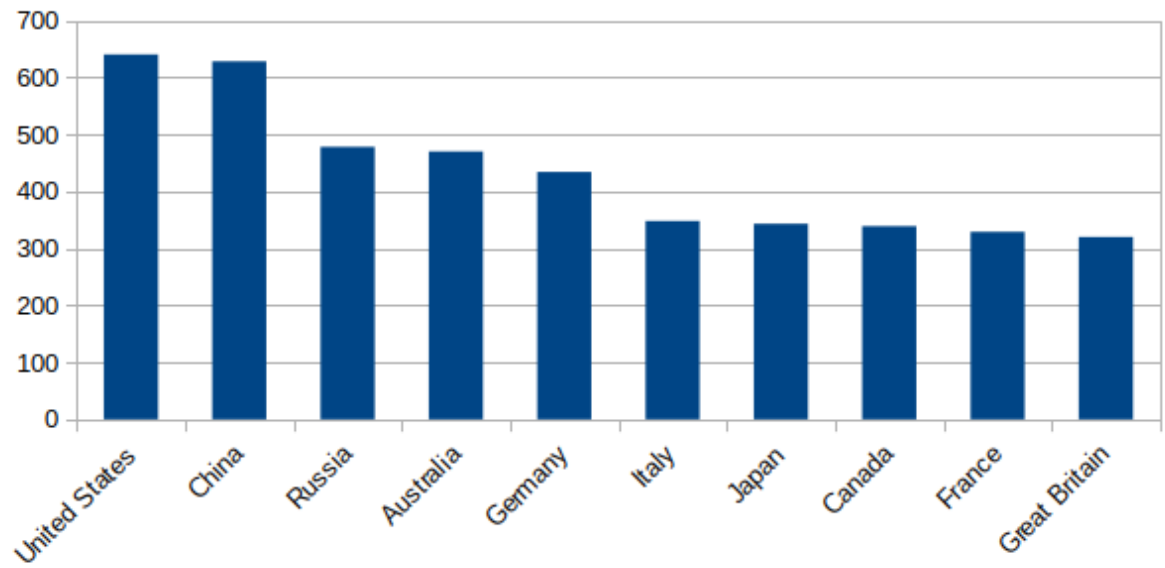
Top 10 2004 (Summer, Greece)



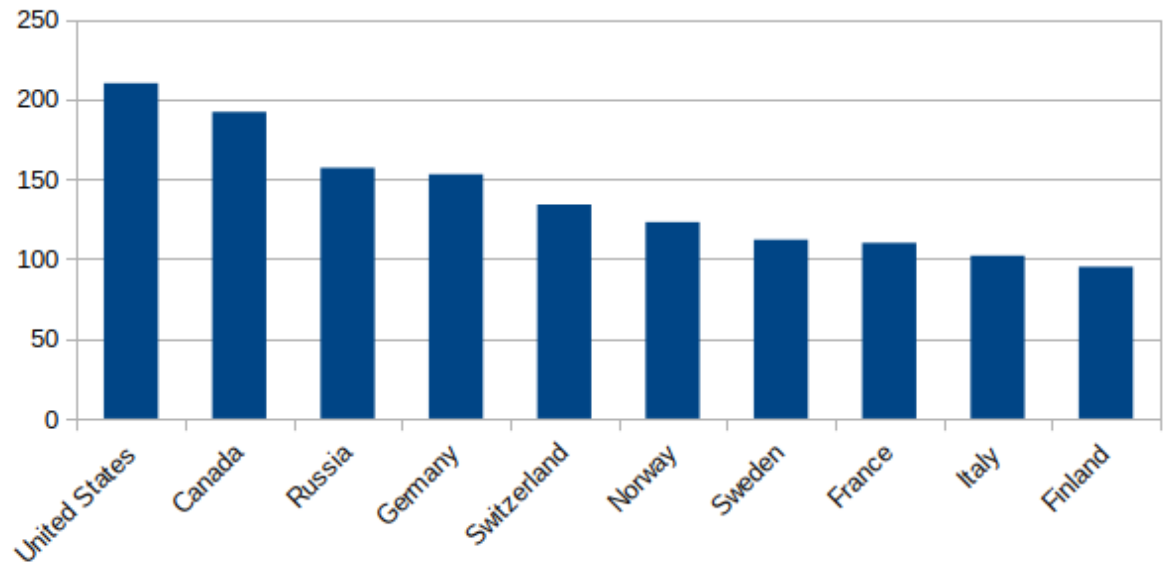
Top 10 2006 (Winter, Italy)



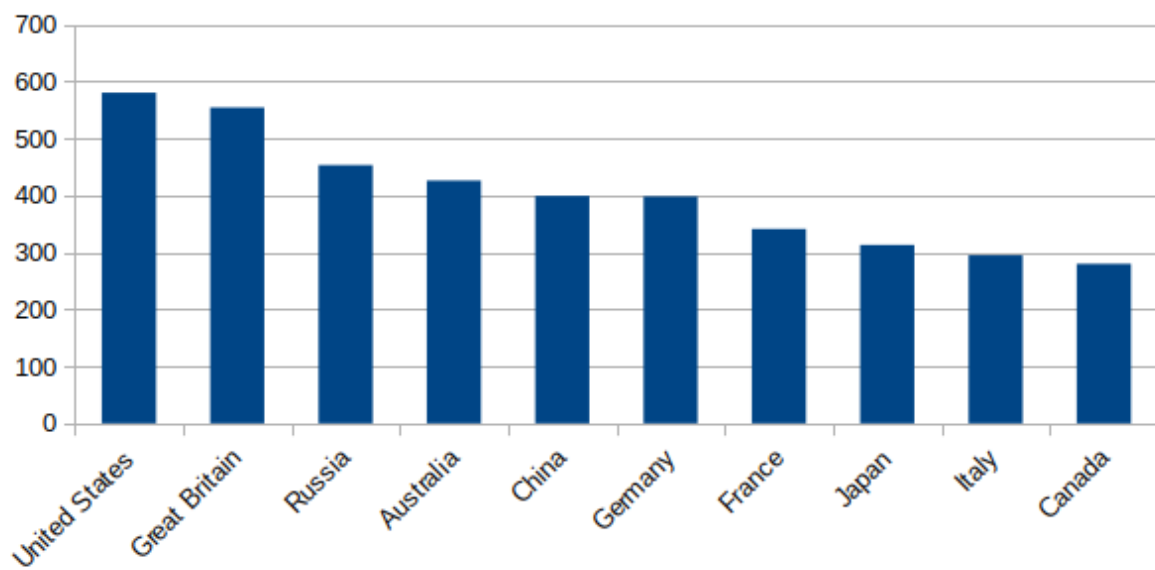
Top 10 2008 (Summer, China)



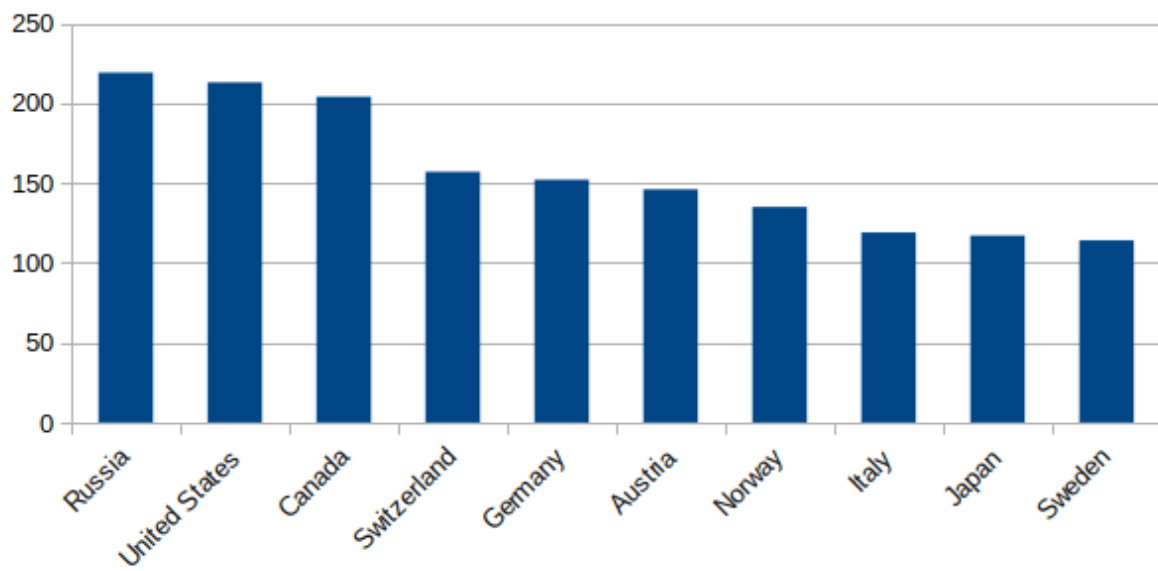
Top 10 2010 (Winter, Canada)

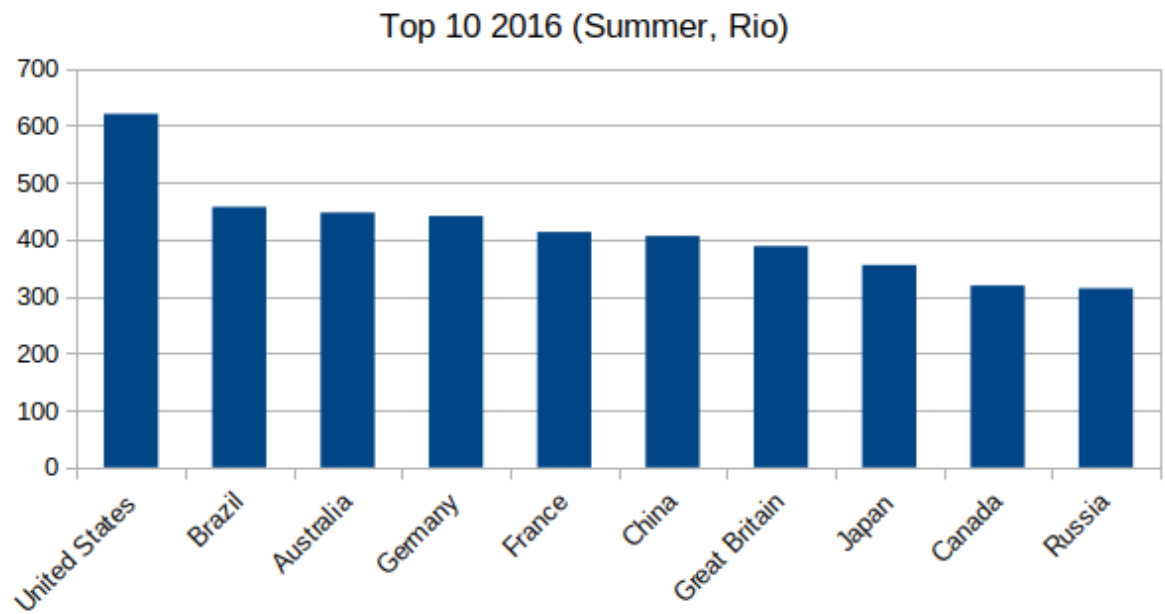


Top 10 2012 (Summer, UK)



Top 10 2014 (Winter, Russia)





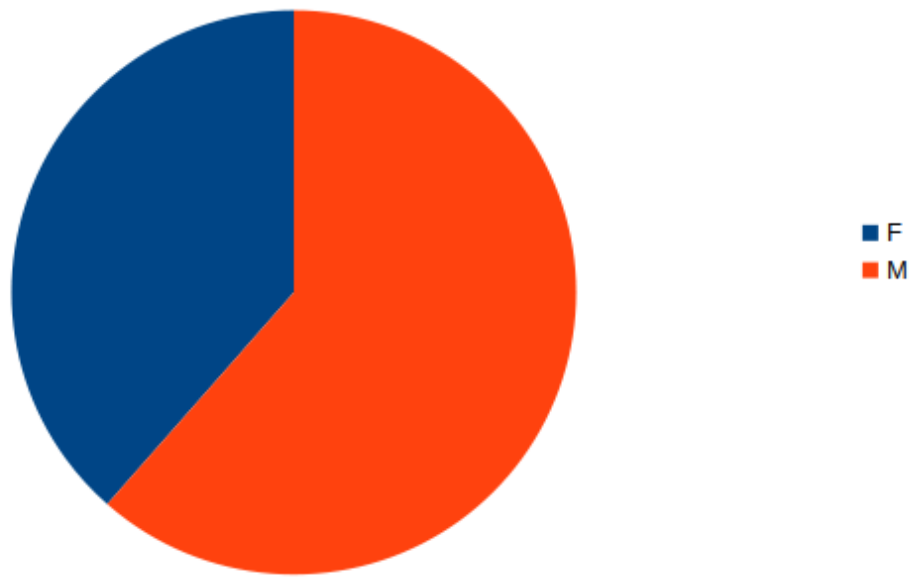
**Answer.**

## Athletes by sex

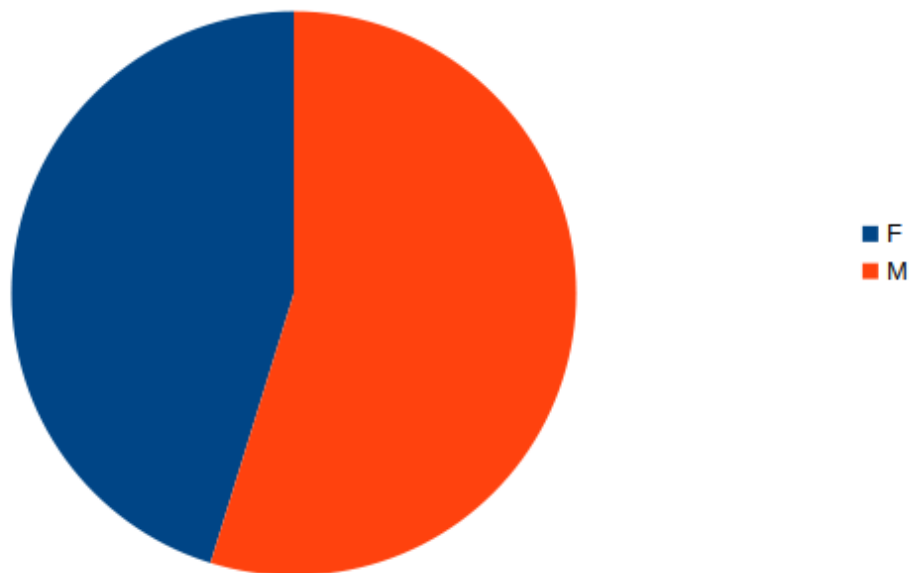
These pie charts show the number of athletes by sex in Sydney 2000 and Rio 2016:



Male and female athletes in Sydney 2000



Male and female athletes in Rio 2016



## Exercise 4

### Exercise 4.1

We want to be more precise. How many male and female athletes were there in Rio 2016 and Sydney 2000?

**Hint:** You can use the **COUNTIFS()** function to solve this exercise. This function works very similarly to the **COUNTA()** function, with the difference that it only counts those cells that meet a certain condition. Feel free to look this function up on the Internet!

**Instructions:**

- Write your formulas in cell AB159 , AB161 , AB166 , and AB168 in the ANSWER tab.

**Answer.**

### Exercise 4.2

Use Excel to calculate the ratio of  $\frac{male}{female}$  athletes in both Rio and Sydney. Has it changed?

**Instructions:**

- Write your formulas in cell AB210 , and AB212 in the ANSWER tab.
- Write your written analysis in cell AR208 in the ANSWER tab.

**Answer.**

### 4.3

Complete the table in the sport\_by\_sex tab.

**Hint:** Use the **COUNTIFS()** function. It works like **COUNTIF()** but allows you to have more than one condition over more than one column.

**Instructions:**

- Complete Column C in the sport\_by\_sex
- Complete Column D in the sport\_by\_sex
- Complete Column E in the sport\_by\_sex

### Exercise 4.4

Interpret the results of your completed table from Exercise 4.3. What can you say?

**Instructions:**

- Write your written analysis in cell X289 in the ANSWER tab.

- Respond with two observations from the data and previous calculations made. Responses should be backed with evidence from the data. Write at least three sentences that are clear and concise, with few grammatical and spelling errors.

**Answer.**

## Medals

### Exercise 5

#### Exercise 5.1

How many medals were awarded between 1998 and 2016, inclusive?

**Hint:** Use the `Won a medal?` column.

**Instructions:**

- Write your formulas in cell `X318` in the `ANSWER` tab.

**Answer.**

#### Exercise 5.2

What is the average number of medals won per athlete in Rio 2016?

**Instructions:**

- Write your formulas in cell `X352` in the `ANSWER` tab.

**Answer.**

#### Exercise 5.3

What is the average number of medals won per athlete in Sydney in 2000? Is this higher or lower than the average in Rio 2016?

**Instruction:**

- Write your formulas in cell `X386` in the `ANSWER` tab.
- Write your written analysis in cell `AR384` in the `ANSWER` tab.

**Answer.**

### Exercise 6

#### Exercise 6.1

Complete the table in the `medals_sport` worksheet to find the top 10 sports with the most medals per athlete in Rio 2016 and Sydney 2000.

**Instruction:**

- Complete Column C in the `medals_sport` worksheet
- Complete Column D in the `medals_sport` worksheet
- Complete Column E in the `medals_sport` worksheet
- Sort the data to find the top 10 sports in 2016 and in 2000.

**Hint 1:** To find the top tens, you will need to use custom sort to sort multiple columns. Namely by year followed by medals per athlete. If you need help, don't hesitate to ask your LI or TA.

**Hint 2:** You can use `COUNTIF()` and `SUMIF()`.

**Hint 3:** Your sorted chart will appear automatically in the `ANSWER` sheet.

**Answer.**

## Exercise 6.2

Which sports are included in both rankings? What could be the reason that these sports show up in both tables?

**Instructions:**

- Compare both sorted tables to find common sports on the lists.
- Provide reasons why these sports appear in both tables.
- List the sports that appeared in both Tables in cell `X469` in the `ANSWER` tab.
- Write your written analysis in cell `X473` in the `ANSWER` tab.

**Answer.**

## Attribution

"120 years of Olympic history: athletes and results", June 15, 2018, Kaggle (user rgriffin, with data from [www.sports-reference.com](http://www.sports-reference.com)), Sports Reference Terms of Use (<https://www.sports-reference.com/termsfuse.html>), <https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results> (<https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results>)