

# DATA301 - Assignment 2 - Data Visualisation

05/08/2024

## Introduction

For this assignment, you will be creating a simple interactive visualisation of Summer Olympic Medal results using Shiny in RStudio.

The interactive visualisation will display data that has been stored in CSV files courtesy of Kaggle user RGriffen, who has made the data available via Kaggle under a Public Domain CC0 1.0 Universal Deed.

The interactive aspect of the visualisation will be that the user will be allowed to filter the data by different features, such as by particular sport, national olympic committee (NOC) or sex, and the Application will automatically update the graphs based on the user's selections.

There is one large CSV file included in the online data set.

## Description of Data

The athlete\_events.csv file covers the results of the Modern Olympics up until the 2016 Rio de Janeiro Olympics.

Each record relates to a single athlete in a single event at a single games. Note that athletes who appear in multiple events in the same or multiple games will have multiple rows.

**Note that the file contains both Summer and Winter Olympics. For this assignment, we are only using the Summer Olympics**

- An ID which is unique to each athlete. The ID is the same for each athlete even if they have played in multiple Games.
- Athlete's name.
- Athlete's sex. (two values are present: 'F' and 'M')
- Athlete's age at the time of the Games for that row.
- Athlete's height in cm. (some of these are NA)
- Athlete's weight in kg. (some of these are NA)
- Athlete's team.
- The National Olympic Committee (NOC) the athlete is representing.
- The name of the Games (e.g. "1992 Summer", "2000 Summer")
- The year the Games were played in.
- The season of the Games. (two values are present: "Summer" and "Winter")
- The host city.
- The sport for this specific row.
- The event for this specific row. A sport will typically have multiple events. For example, the sport of "athletics" has events such as, amongst many others, "Athletics Women's 100 metres" and "Athletics Women's 4 x 100 metres relay".
- The result for that athlete in that sport in that event for that Games. Most rows are "NA", however "Gold", "Silver" or "Bronze" shows where the athlete in this row in this sport in this event in this Games medalled.

You may not need all of the data in the CSV file.

### Q1 (6 marks)

a) [3 marks] Create a Shiny App in RStudio that can read in the *athlete\_events.csv* file and displays a stacked bar chart showing the total number of gold medals, silver medals and bronze medals won by each NOC. NOCs will be plotted on the x-axis and medal counts on the y-axis. **Note: only one medal per event should be counted, so a team of four rowers who each win a gold medal in a single event only counts as one gold medal for that country.**

b) [3 marks] Add a second chart - a scatter plot - that displays the number of athletes vs the number of events across all games. The x-axis should be the number of athletes who competed at a particular games, and the y-axis should be number of events at that particular games. An individual point in the scatter plot represents a single games (e.g. one for Sydney 2000, one for Paris 1924, etc)

### Q2 (8 marks)

a) [1 mark] Modify your application so the graph in Q1(a) can be toggled between stacked bar chart mode and side-by-side bar chart mode.

b) [1 mark] Modify your application so that the user may select a specific set of sports and the graphs created in Q1(a) and Q1(b) are automatically updated to reflect results only from those selected sports. **Note that most sports have multiple events, such as swimming and athletics.**

c) [2 marks] Modify your application so that the user may elect to show only specific NOCs (a set of NOCs may be selected, rather than just one NOC) and a specific year range (using a sliding widget). The graphs created in Q1(a) and Q1(b) are automatically updated to reflect only those selected NOCs and years.

d) [4 marks] [CHALLENGE] Modify your graphs from Q1(a) and Q1(b) by allowing the user to directly compare side-by-side the number of male and female athletes and the number of male and female medal winners. The graphs should also respond to the filters created in Q2(a) – Q2(c). Ensure there are appropriate visual cues (such as, for example, labels and titles) for the user to understand what they are looking at.

### Q3 (6 marks)

Consider the visualisation at: [Disney Film Dialogue Interactive Visualisation](#)

As a user of the visualisation, briefly describe:

a) [3 marks] one aspect of the visualisation that is a problem with respect to either the *gulf of execution* or the *gulf of evaluation*

b) [3 marks] one aspect of the visualisation that is a positive benefit with respect to either the *gulf of execution* or the *gulf of evaluation*

[Note: in your answers, please identify what the aspect is, which gulf you are referring to, and a couple of sentences on why it is a problem or a benefit.]

### Submission

Please note that you should only submit one version of your Application: the one which satisfies the most number of questions above. Your answers to Question 1 and 2 can be submitted simply as an App.R file.

**Important: Please make sure that when you do not change the name of the datafile provided in your code and please make sure that the App.R file is written to assume the data file is in the same directory as the App.R file.**

**DO NOT hardcode the datafile name in your R file to be on your own account / device**

**DO NOT submit the csv file as it is 40+ MB.**

Your answers to Question 3 can be submitted either as a PDF document or simply as an R Markdown document.